

# Resource-Constrained Adaptive Search and Tracking for Sparse Dynamic Targets

Gregory Newstadt, *Member, IEEE*, Dennis Wei, *Member, IEEE*, and Alfred O. Hero III, *Fellow, IEEE*

April 9, 2014

## Abstract

This paper considers the problem of resource-constrained and noise-limited localization and estimation of dynamic targets that are sparsely distributed over a large area. We generalize an existing framework [Bashan et al, 2008] for adaptive allocation of sensing resources to the dynamic case, accounting for time-varying target behavior such as transitions to neighboring cells and varying amplitudes over a potentially long time horizon. The proposed adaptive sensing policy is driven by minimization of a surrogate function for mean squared error within locations containing targets. We provide theoretical upper bounds on the performance of adaptive sensing policies by analyzing solutions with oracle knowledge of target locations, gaining insight into the effect of target motion and amplitude variation as well as sparsity. Exact minimization of the multi-stage objective function is infeasible, but myopic optimization yields a closed-form solution. We propose a simple non-myopic extension, the Dynamic Adaptive Resource Allocation Policy (D-ARAP), that allocates a fraction of resources for exploring all locations rather than solely exploiting the current belief state. Our numerical studies indicate that D-ARAP has the following advantages: (a) it is more robust than the myopic policy to noise, missing data, and model mismatch; (b) it performs comparably to well-known approximate dynamic programming solutions but

Gregory Newstadt and Alfred Hero are with the Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, E-mail: ({newstage},{hero}@umich.edu).

Dennis Wei is with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA, E-mail: dwei@us.ibm.com.

The research in this paper was partially supported by Air Force Office of Scientific Research award FA9550-06-1-0324, by Air Force Research Laboratory award FA8650-07-D-1221-TO1, and by Army Research Office MURI grant number W911NF-11-1-0391.

This work was presented in part at IEEE CAMSAP 2013 and Asilomar Conference on Signals, Systems and Computers 2011.

at significantly lower computational complexity; and (c) it improves greatly upon non-adaptive uniform resource allocation in terms of estimation error and probability of detection.

## I. INTRODUCTION

Systems for wide area surveillance such as the Gotcha synthetic aperture radar and the Angel Fire electro-optical sensor currently offer near real-time imaging and surveillance of city-sized scenes. Generally, these systems perform continuous collection of data, followed by forensic analysis to detect and track targets within the scene. The amount of raw data collected by these systems is often very large, often collected uniformly over a large area, even though the interesting features (e.g. moving targets, etc.) exist only in a few locations. Due to such data collection inefficiencies, there is a high likelihood that much sensing and/or computational resources may be wasted by searching areas where targets are not located. Abidi et al [1] provides a comprehensive survey of recent work in wide-area surveillance, including topics in coverage analysis, optimal sensor positioning and sensor fusion. They stress the need for intelligent data collection and system analysis in order to deal with such data collection inefficiencies.

An alternative is adaptive sampling, for which past observations are used to inform the collection of future observations, with the goal of focusing effort onto the “interesting” regions of the search space. Adaptive sampling can be an important tool for efficiently managing the collection inefficiency problem faced by wide-area surveillance systems. Previous work has shown that, when constrained to use equal resources, adaptive sampling can significantly improve target localization performance [2]–[11] in comparison to a uniform policy that uses equal sensing effort across the scene. Benefits of adaptive sensing include: gains in estimation precision [4], [5]; provable detection of the targets often at faster convergence rates [7], [10]; and improved robustness in detection performance as measured by the minimum detectable amplitudes [8].

Bashan et al [4] provided an optimal two-stage policy called ARAP for adaptively localizing targets and estimating their amplitudes in noise under effort budget constraints. A multiscale approach was subsequently introduced in order to further reduce the total number of measurements [5]. Hitchings and Castanon [6] provided an online modification to [4] through Lagrangian constraint relaxation. Additionally, [9] extended the two-stage policy to an arbitrary number of stages using approximate dynamic programming, and generalized the framework to allow for a variety of measurement/estimation loss models. Using a similar model called Distilled Sensing, [7], [8] specified a methodology for locating targets in noise at much lower signal-to-noise ratios (SNR) than non-adaptive methods. Malloy et al [10] extends Distilled Sensing to non-Gaussian models, while [8], [11] consider compressive measurements.

In the above methods, it is assumed that the targets of interest remain stationary across sensing/observation epochs. In wide-area surveillance, however, many targets exhibit complex dynamic behavior such as movement, entering/leaving the scene, and obscuration. Krishnamurthy [12] considers the problem of selecting the direction to point an agile sensor in order to track  $P$  moving targets among a finite number of cells. When the state is fully observable, the problem can be posed as a Markov decision process (MDP). Krishnamurthy formulates the problem as a hidden Markov model (HMM) tracking problem in the more challenging case where the state is observed with noise. He discusses an optimal policy which depends on the individual target's belief state - the conditional density of the state given the observation history. Moreover, a suboptimal approach to approximating the optimal selection criterion is provided to combat the prohibitive computational complexity of the optimal solution.

Chong et al. [13] show that many adaptive sensing problems can be formulated as partially observable Markov decision processes (POMDPs). This general framework is concerned with selecting actions to maximize the sum of an immediate reward and an expected future reward, given a system with Markovian evolution but only partially observable state (e.g. due to noise). Unfortunately, the expected future reward is often difficult to compute, while the tractability of the optimal solution is further impaired when the state space and action space are large. Chong et al. [13] propose approximate methods that include parametric approximations, reinforcement learning techniques, and rollout policies. The last of these assumes that a base policy is available that may not be optimal but is simple to compute. Rollout policies then ensure policy improvement, i.e., they are guaranteed to do at least as well as the base policy. However, rollout policies still remain impractical in cases where the action space is large.

In this paper, we extend the Bayesian formulation of adaptive sampling [4] to targets that are dynamic as well as sparsely located in the scene. Our model encompasses target motion, appearance and disappearance, and target amplitude variation. This formulation can simultaneously account for multiple targets as well as allocation of continuous-valued sensing resources, such as energy, in contrast to the discrete resource allocation formulation in [12]. Our formulation is based on a simple approximation to the target posterior distribution and a multistage extension of the cost function in [4]. In the context of this framework we then introduce the Dynamic Adaptive Resource Allocation Policy (D-ARAP), a non-myopic policy for adaptive sampling that achieves a favorable trade-off between performance and planning complexity. Our analysis suggests that as compared to approximate POMDP solutions, in particular rollout policies, D-ARAP performs well but at a fraction of the computational cost. Compared to myopic policies, D-ARAP has increased robustness to noise, missing data, and model mismatch. Lastly, compared to the non-adaptive uniform policy, D-ARAP continues to yield large improvements in estimation and detection

performance similar to the static case [4], [9].

For static targets, it is known [4], [5], [14] that estimation gains due to adaptive sampling increase in the sparse target regime, where targets occupy few locations in the search area. The present work shows that the benefits of adaptive sensing framework [4] can be extended to the dynamic setting. We derive upper bounds on adaptive performance gains through analysis of omniscient and semi-omniscient policies with complete or partial knowledge of target locations over time, respectively. The bounds confirm the benefit of the sparse target regime for dynamic targets. They also characterize the effect of target motion and amplitude variation on the potential gain. We show simulations that indicate that D-ARAP can approach the semi-omniscient performance as the SNR and number of stages increase. Furthermore, comparison of the omniscient and semi-omniscient policies allows us to quantify the effect of partial and causal knowledge of target motion.

The rest of this paper is organized as follows. We formalize the problem in Section II and present adaptive sensing policies in Section III. We derive performance bounds for adaptive sensing of dynamic targets in Section IV. Numerical performance analysis is given in Section V. In Section VI, we conclude and discuss future directions.

## II. PROBLEM FORMULATION

We consider a space  $\mathcal{X} = \{1, 2, \dots, Q\}$  containing  $Q$  cells and a time-varying region of interest (ROI)  $\Psi(t) \subset \mathcal{X}$ ,  $t = 1, \dots, T$ . Let  $i$  be a location in  $\mathcal{X}$  and define  $I_i(t)$  to be the indicator that  $i$  is in the ROI at time  $t$ , i.e.,  $I_i(t) = 1$  if  $i \in \Psi(t)$  and  $I_i(t) = 0$  otherwise. We use a probabilistic target model in which  $I_i(1) = 1$  with prior probability  $p_i(1)$ , independently of the other indicators. For  $I_i(t) = 0$ , the corresponding signal amplitude  $\theta_i(t)$  is taken to be zero, while for  $I_i(t) = 1$ , the amplitude  $\theta_i(t)$  is modeled as a Gaussian random variable. The initial amplitudes  $\theta_i(1)$ ,  $i \in \Psi(1)$  are drawn independently with means  $\mu_i(1)$  and variances  $\sigma_i^2(1)$ . As in previous work [4], [9], a non-informative uniform prior on target locations/amplitudes is assumed with  $p_i(1) = p_0$ ,  $\mu_i(1) = \mu_0$  and  $\sigma_i^2(1) = \sigma_0^2$  for all  $i$ , although non-uniform priors could also be accommodated.

We generalize previous work [4], [9] by introducing a dynamic target state model with state transitions and a birth-death model for target appearance/disappearance. To describe the model, we index the targets by target number instead of by cell: Let  $s^{(n)}(t) \in \Psi(t)$ ,  $n = 1, \dots, |\Psi(t)|$  be the position of the  $n$ -th target at time  $t$  and  $v^{(n)}(t) = \theta_{s^{(n)}(t)}(t)$  be its associated amplitude. Let  $\alpha$  be the probability that each target is removed from the scene at each time. The target transition model and amplitude update for

remaining targets is

$$\Pr(s^{(n)}(t+1) = i | s^{(n)}(t) = j) = \begin{cases} (1-\alpha)\pi_0, & i = j \\ \frac{(1-\alpha)(1-\pi_0)}{|G(j)|}, & i \in G(j) \end{cases}, \quad (1)$$

$$\vartheta^{(n)}(t+1) = \vartheta^{(n)}(t) + Z^{(n)}(t) \quad (2)$$

where  $\pi_0$  is the probability that a target remains in the same location,  $G(j)$  is the set of cells that are neighbors of cell  $j$ , and  $Z^{(n)}(t)$  is a zero mean white Gaussian noise with variance  $\Delta^2$ :

$$Z^{(n)}(t) \sim \mathcal{N}(0, \Delta^2). \quad (3)$$

$\Delta^2 > 0$  captures the variance of random perturbations to the target amplitudes. The model (2) can be used to approximate the effect of model mismatch, target fluctuations, or scattering of the radar signal. In each of these cases, measurements at each stage are discounted by the increase in uncertainty due to the error sources.

Let  $B(t)$  be the event that a single new target enters the scene at time  $t$  with probability  $\beta$ . Then conditioned on  $B(t)$ ,

$$\begin{aligned} s^{(|\Psi(t)|+1)}(t+1) | B(t) &\sim \text{Uniform}\{1, 2, \dots, Q\}, \\ \vartheta^{(|\Psi(t)|+1)}(t+1) | B(t) &\sim \mathcal{N}(\mu_0, \sigma_0^2). \end{aligned} \quad (4)$$

We restrict our attention to the case where at most one target occupies a cell at any instant. In the sparse situations considered here (i.e.  $p_0 \ll 1$ ), this occurs with high probability.

Observations are made in  $T$  stages with effort levels  $\lambda_i(t)$  that vary with location  $i$  and time  $t$ . In general, effort might be computing power, complexity, cost, or energy that is allocated to probing a particular cell. It is assumed that the quality of an observation increases with effort. Given  $\lambda_i(t)$ , the corresponding observation  $y_i(t)$  takes the form

$$y_i(t) = \sqrt{\lambda_i(t)} I_i(t) \theta_i(t) + n_i(t), \quad t = 1, \dots, T, \quad (5)$$

where  $n_i(t)$  represents i.i.d. zero-mean Gaussian noise with variance  $\sigma^2$ . The total effort in each stage is constrained as  $\sum_{i=1}^Q \lambda_i(t) \leq \Lambda(t)$ .

The goal is to estimate the target state  $\left\{ \Psi(t), \{\theta_i(t)\}_{i \in \Psi(t)} \right\}_{t=1, \dots, T}$  over  $T$  stages. The posterior distribution of the target state conditioned on measurements is therefore of central interest. When the targets are static, the posterior distribution factors by cell and can be exactly represented by the posterior

mean/variance of the target amplitude and the posterior probability of target existence in each cell. In the dynamic case, there is no simple factorization that allows for efficient exact estimation of the posterior distribution, partly due to the fact that the posterior distribution of the amplitudes becomes a Gaussian mixture (due to nonzero transition probabilities to neighboring cells) rather than a simple univariate Gaussian. As an alternative, the posterior distribution may be approximated using several standard approaches including particle filters, extended Kalman filters, and Unscented Kalman filters, with varying tradeoffs between accuracy and computational burden.

Here we propose a simple approximation to the posterior that is accurate under two conditions: (a) at most one target occupies the vicinity of a cell at any one time; and (b) the Gaussian mixture is well represented by the most likely Gaussian mixture component; i.e., the component belonging to the most probable trajectory of the target given the measurements. Further details on this approximation are available in Chapter 3 of [15]. The first condition is valid when there is very low probability that targets will cross tracks. In practice, one could relax this condition by using methods such as the Joint Multitarget Probability Density Filter (JMPD) [16], which independently tracks targets when they are far apart, while jointly tracking targets that are close to each other. However, we do not address this generalization in this paper. The second condition is equivalent to the existence of a dominant mode in the Gaussian mixture characterizing the posterior density.

Under conditions (a) and (b), the posterior distribution can be approximated with the following:

$$p_i(t) = \Pr(I_i(t) = 1 | \mathbf{Y}(t-1)), \quad (6)$$

$$\mu_i(t) = \mathbb{E}[\theta_i(t) | I_i(t) = 1, \mathbf{Y}(t-1)], \quad (7)$$

$$\sigma_i^2(t) = \text{var}[\theta_i(t) | I_i(t) = 1, \mathbf{Y}(t-1)]. \quad (8)$$

where  $\mathbf{Y}(T) = \{y_i(t)\}_{i \in \mathcal{X}, t \in \{1, \dots, T\}}$  is the sequence of observations. For brevity, we denote the collection of posterior probabilities, means, and variances as

$$\mathbf{x}(t) = \{p_i(t), \mu_i(t), \sigma_i^2(t)\}_{i=1}^Q \quad (9)$$

This representation may be combined with a particle filter, e.g. one using the JMPD [16], for target state estimation, while using the above model for resource planning.

### III. SEARCH POLICY FOR DYNAMIC TARGETS UNDER RESOURCE CONSTRAINTS

In this section, we provide methods for determining for a sequence of effort allocations  $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}(t)\}_{t=1}^T$  where  $\boldsymbol{\lambda}(t) = \{\lambda_1(t), \dots, \lambda_Q(t)\}$ .  $\boldsymbol{\lambda}(t)$  is a mapping from the previous observations  $\mathbf{Y}(t-1)$  to  $[0, \Lambda(t)]^Q$  and is called the allocation policy.

### A. Optimization objective

The following is a multistage extension of the cost function in [4], [9]:

$$J_T(\boldsymbol{\lambda}) = \mathbb{E} \left[ \sum_{t=1}^T \gamma(t) \sum_{i=1}^Q \frac{p_i(t)}{\sigma^2/\sigma_i^2(t) + \lambda_i(t)} \right], \quad (10)$$

where  $\{\gamma(t)\}_{t=1}^T$  is a set of known weights on different planning stages. The cost function (10) corresponds exactly to the MSE for estimating target amplitudes  $\{\theta_i(t)\}_{i \in \Psi(t)}$  in two cases: (a) when targets are stationary (but amplitudes may vary); and (b) when target locations may change but are known exactly (i.e.  $p_i(t) = I_i(t)$ ). We define the per-stage cost:

$$M_t(\boldsymbol{\lambda}; \mathbf{x}(t)) = \sum_{i=1}^Q \frac{p_i(t)}{\sigma^2/\sigma_i^2(t) + \lambda_i(t)}. \quad (11)$$

Recalling from (6) that  $p_i(t) = \mathbb{E}[I_i(t) = 1 \mid \mathbf{Y}(t-1)]$ , the expected per-stage cost can also be expressed as

$$\mathbb{E} [M_t(\boldsymbol{\lambda}; \mathbf{x}(t))] = \mathbb{E} \left\{ \sum_{i \in \Psi(t)} \frac{1}{\sigma^2/\sigma_i^2(t) + \lambda_i(t)} \right\}, \quad (12)$$

where the expectation is taken over both  $\Psi(t)$  and  $\mathbf{Y}(t-1)$ .

### B. Optimal dynamic programming solution

The optimal effort allocation problem can be stated as

$$\{\hat{\lambda}_i(t)\}_{i,t} = \arg \min_{\boldsymbol{\lambda}} J_T(\boldsymbol{\lambda}), \quad (13)$$

where  $\{\hat{\lambda}_i(t)\}_i$  is a function of  $\mathbf{Y}(t-1)$  and

$$\sum_{i=1}^Q \lambda_i(t) \leq \Lambda(t), \quad t = 1, 2, \dots, T. \quad (14)$$

Dynamic programming (DP) can be used to exactly obtain an optimal policy that minimizes equation (10). In the case when  $\gamma(T) = 1$  and  $\gamma(t) = 0, t < T$ , this policy is given by a sequence of recursive minimizations that proceed as follows<sup>1</sup>

$$K_T(\mathbf{x}(T)) = \min_{\boldsymbol{\lambda}(T)} M_T(\boldsymbol{\lambda}; \mathbf{x}(T)), \quad \sum_{i=1}^Q \lambda_i(T) = \Lambda(T) \quad (15)$$

<sup>1</sup>As a minor technical point, if we consider general values for the weights  $\gamma(t)$ , then equation (16) requires an additional term for the current cost at stage  $t$ .

and define recursively for  $t = T - 1, T - 2, \dots, 1$

$$\begin{aligned} K_t(\mathbf{x}(t)) &= \min_{\boldsymbol{\lambda}(t)} \mathbb{E} \left[ K_{t+1}(\mathbf{x}(t+1)) \middle| \mathbf{x}(t), \boldsymbol{\lambda}(t) \right], \\ \text{s.t.} \quad &\sum_{i=1}^Q \lambda_i(t) = \Lambda(t). \end{aligned} \tag{16}$$

Wei and Hero [9] show that this solution is only tractable for  $T \leq 2$ . This is an artifact of the difficulty in computing the expectation in (16), which is generally approximated with Monte Carlo samples, as well as the fact that  $\boldsymbol{\lambda}(t)$  lies in a multi-dimensional space for  $t > 1$ . For  $T > 2$ , we therefore have to consider approximations to the optimal policy. In the next sections, we provide a myopic solution that optimizes  $M_t(\boldsymbol{\lambda}; \mathbf{x}(t))$  for  $t = 1, 2, \dots, T$  without recursion (i.e., assuming that  $t$  is the last stage) and an alternative policy that improves upon the myopic solution with low additional computational cost.

### C. Myopic policy

The myopic optimization problem at time  $t$  is given by

$$\min_{\boldsymbol{\lambda}(t)} M_t(\boldsymbol{\lambda}(t); \mathbf{x}(t)) \quad \text{s.t.} \quad \sum_{i=1}^Q \lambda_i(t) = \Lambda(t) \tag{17}$$

where  $\boldsymbol{\lambda}(t)$  depends on previous observations  $\mathbf{Y}(t-1)$  through  $\mathbf{x}(t)$ . The optimal solution, similar to the one given in [9], begins by defining  $\chi$  to be an index permutation that sorts the quantities  $\sqrt{p_i(t)}\sigma_i^2(t)$  in non-increasing rank order:

$$\sqrt{p_{\chi(1)}(t)}\sigma_{\chi(1)}^2(t) \geq \dots \geq \sqrt{p_{\chi(Q)}(t)}\sigma_{\chi(Q)}^2(t). \tag{18}$$

Let  $c_i(t) = \sigma^2/\sigma_i^2(t)$ . Then define  $g(k)$  to be the monotonically non-decreasing function of  $k = 0, 1, \dots, Q$  with  $g(0) = 0$ ,  $g(Q) = \infty$ , and

$$g(k) = \frac{c_{\chi(k+1)}(t)}{\sqrt{p_{\chi(k+1)}(t)}} \sum_{i=1}^k \sqrt{p_{\chi(i)}(t)} - \sum_{i=1}^k c_{\chi(i)}(t) \tag{19}$$

for  $k = 1, \dots, Q - 1$ . Then the solution to (17) is

$$\lambda_{\chi(i)}^m(t) = \left( \Lambda(t) + \sum_{j=1}^{k^*} c_{\chi(j)}(t) \right) \frac{\sqrt{p_{\chi(i)}(t)}}{\sum_{j=1}^{k^*} \sqrt{p_{\chi(j)}(t)}} - c_{\chi(i)}(t), \tag{20}$$

for  $i = 1, \dots, k^*$  and  $\lambda_{\chi(i)}^m(t) = 0$  for  $i = k^* + 1, \dots, Q$ . The number of nonzero components  $k^*$  is determined by the interval  $(g(k-1), g(k)]$  to which the budget parameter  $\Lambda(t)$  belongs. Since  $g(k)$  is monotonic, the mapping from  $\Lambda(t)$  to  $k^*$  is well-defined.



#### D. Non-myopic extension

We propose a simple improvement to the myopic policy that combines exploitation of the current belief state and exploration of the scene at large. The proposed non-myopic allocation policy is called the Dynamic Adaptive Resource Allocation Policy, or D-ARAP, and is defined by

$$\lambda_i^d(t; \kappa(t)) = [\kappa(t)]\lambda^u(t) + [1 - \kappa(t)]\lambda_i^m(t), \quad (21)$$

where  $\kappa(t) \in [0, 1]$  is the exploration coefficient,  $\lambda^u(t) = \Lambda(t)/Q$  is the uniform allocation policy, and  $\lambda_i^m(t)$  is given by (20). Note that the first term in (21) allocates a percentage of the resources uniformly to the scene, while the second term weights allocations according to the myopic solution (20).

We define the full set of exploration coefficients for a  $T$ -stage policy as  $\kappa(T) = \{\kappa_T(t)\}_{t=1}^T$ , where the subscript  $T$  indicates the number of stages when needed for clarity. In the rest of the paper, we often use vector notation to represent the allocations to all  $T$  stages, defined as

$$\lambda^d(\kappa(T)) = \left\{ \lambda_i^d(t; \kappa(t)) \right\}_{i=1, \dots, Q, t=1, \dots, T}. \quad (22)$$

Without prior knowledge on the location of targets, the first stage should be purely exploratory, i.e.,  $\kappa_T(1) = 1$ . In addition, since the last stage should be purely exploitative or myopic, we set  $\kappa_T(T) = 0$ . To determine  $\kappa(T)$ , we consider both offline policies, which are determined prior to collecting observations, and online policies, which are determined adaptively as measurements are collected. Note that  $\lambda_i^m(t)$  is a function of previous measurements. Thus, the offline policies can still be data-dependent as long as  $\kappa_T(t) < 1$ .

#### E. Rollout policy

We first describe an offline policy called the “offline rollout policy” that is recursive in the sense that a  $T$ -stage policy is created by building upon a previously defined  $(T - 1)$ -stage policy. This method also requires a “base” policy,  $\phi(T_0)$  which pre-defines the last  $T_0$  stages of a  $T$ -stage policy. Rollout policies for general dynamic programming problems are discussed in great detail in [17]. In this context, the simplest rollout policy is just  $\phi(1) = \{0\}$ , which indicates that the last stage should be purely exploitative. The pseudocode for the offline rollout policy is given in Fig. 1 and yields policies for  $\{\kappa(\tau)\}_{\tau=T_0+1}^T$  from  $(T_0 + 1)$  to  $T$  inclusive. Define  $\omega_{\tau-1}(t) = \{\kappa_{\tau-1}(t')\}_{t'=1}^t$  to be the first  $t$  values of the previous policy  $\kappa(\tau - 1)$ . Then in each iteration, a  $\tau$ -stage policy is constructed as  $\kappa(\tau) = \{\omega_{\tau-1}(\tau - 1 - T_0), \kappa(\tau - T_0), \phi(T_0)\}$  where  $\kappa(\tau - T_0)$  is a single parameter which we search over. The values of  $\kappa(\tau - T_0)$  are chosen to minimize the full non-myopic cost in (10).

```

procedure  $\{\kappa(\tau)\}_{\tau=T_0+1}^T = \text{OfflineRolloutPolicy}(\phi(T_0))$ 
  Set  $\omega(1) = 1, \kappa(T_0 + 1) = \{\omega(1), \phi(T_0)\}$ .
  for  $\tau = T_0 + 2, \dots, T$  do
    for each  $\kappa(\tau - T_0) \in (0, 1]$  do
      Set  $\tilde{\kappa}(\tau) = \{\omega_{\tau-1}(\tau - 1 - T_0), \kappa(\tau - T_0), \phi(T_0)\}$ .
      Calculate  $C(\kappa(\tau - T_0)) = J_\tau(\lambda^d(\tilde{\kappa}(\tau)))$ .
    end for
    Choose  $\hat{\kappa}(\tau - T_0) = \arg \min_{\kappa(\tau - T_0)} C(\kappa(\tau - T_0))$ .
    Set  $\omega_\tau(\tau - T_0) = \{\omega_{\tau-1}(\tau - 1 - T_0), \hat{\kappa}(\tau - T_0)\}$ .
    Set  $\kappa(\tau) = \{\omega_\tau(\tau - T_0), \phi(T_0)\}$ .
  end for
  Return  $\{\kappa(\tau)\}_{\tau=T_0+1}^T$ .
end procedure

```

Fig. 1. Offline rollout policy pseudocode for determining exploration parameters  $\kappa$

The expectation in (10) is approximated with Monte Carlo samples from the belief state  $\mathbf{x}(t)$  for  $t = 1, 2, \dots, \tau$ . This process can be done efficiently by noting that the first  $\tau - T_0 - 1$  stages remain the same for  $\kappa(\tau - 1)$  and  $\kappa(\tau)$ . Therefore, we only need to draw samples for the last  $T_0 + 1$  stages at each iteration, as well as perform a line search over the single parameter  $\kappa(\tau - T_0)$ . Thus, the offline rollout policy requires  $\mathcal{O}(TT_0)$  Monte Carlo simulations to determine policies for  $\{\kappa(\tau)\}_{\tau=T_0+1}^T$ . This improves upon the approach [18] where a nested optimization procedure (aka, the “nested policy”) required  $\mathcal{O}(T^2)$  calculations. In our experiments (not shown), the offline rollout policy performed just as well as the nested policy, though with reduced computational complexity. We do not further discuss the nested policy.

#### F. Myopic+ policy

To further reduce the computational burden, we consider another policy which we call the “myopic+ policy” which requires only  $\mathcal{O}(T)$  expectations to be calculated (once again through Monte Carlo approximation.) Similar to the offline rollout policy, this policy is built in a sequential fashion. Whereas the  $\tau$ -stage offline rollout policy iteratively optimizes over  $\kappa_\tau(\tau - T_0)$  followed by a  $T_0$ -stage base policy, the myopic+ policy chooses  $\kappa_\tau(\tau)$  directly without any subsequent rollout. In particular, we define  $\tilde{\kappa}(\tau) = \{\tilde{\kappa}(\tau - 1), \kappa_\tau(\tau)\}$ . Note that given  $\tilde{\kappa}(\tau - 1)$ , the current state  $\mathbf{x}(\tau)$  is random only

through the noisy measurements  $\mathbf{Y}(\tau - 1)$ . Additionally, given  $\tilde{\kappa}(\tau - 1)$ , minimization of  $J_\tau(\boldsymbol{\lambda}^d(\tilde{\kappa}(\tau)))$  is equivalent to minimization of the following over the single exploration coefficient  $\kappa_\tau(\tau)$ :

$$\mathbb{E}_{\mathbf{Y}(\tau-1)} \left\{ M_\tau(\boldsymbol{\lambda}^d(\{\tilde{\kappa}(\tau - 1), \kappa_\tau(\tau)\}); \mathbf{x}(\tau)) \right\} \quad (23)$$

Note that the quantity within the expectation is always minimized by  $\kappa_\tau(\tau) = 0$ , since, by definition, this value optimizes the myopic cost. To promote exploration, i.e.  $\kappa_\tau(\tau) > 0$ , we adopt a  $(1 + \rho)$ -optimality criterion:

$$\begin{aligned} \hat{\kappa}(\tau) = \\ \max_{\kappa} \left\{ \kappa : B_\tau^D(\tilde{\kappa}(\tau - 1), \kappa) \leq (1 + \rho) B_\tau^D(\tilde{\kappa}(\tau - 1), 0) \right\}, \end{aligned} \quad (24)$$

where  $\rho > 0$  is a tolerance and

$$B_\tau^D(\tilde{\kappa}(\tau - 1), \kappa) = \mathbb{E}_{\mathbf{Y}(\tau-1)} \left\{ M_\tau \left( \boldsymbol{\lambda}^d(\{\tilde{\kappa}(\tau - 1), \kappa\}); \mathbf{x}(\tau) \right) \right\}, \quad (25)$$

where  $\mathbf{x}(\tau)$  is a function of  $\tilde{\kappa}(\tau - 1)$  through the measurements  $\mathbf{Y}(\tau - 1)$ . Since  $\kappa_\tau(\tau) = 0$  optimizes the last-stage cost by definition, (24) results in a policy that is within  $(1 + \rho)$  of the expected minimum myopic cost at each stage. Observe that (23)-(25) are used to build a  $T$ -stage policy in an iterative fashion and have computational complexity  $\mathcal{O}(T)$ . The iterative process is repeated for  $\tau = 2, 3, \dots, T - 1$ , but the final stage is given by the analytical solution  $\kappa_T(T) = 0$ . Pseudocode for the myopic+ policy is given in Fig. 2.

To understand the optimality-criterion in equation (24), it is illustrative to look at Fig. 3 which plots  $B_2^D(1, \kappa)$  as a function of  $\kappa$  for low, medium, and high values of  $\Lambda(2)$  in (a), (b), and (c), respectively. It is seen that in all cases, the myopic cost is optimized when  $\kappa(2) = 0$ . To encourage exploration, (24) increases  $\kappa(2)$ . The amount of increase becomes larger as the SNR decreases. The red dotted line shows a deviation of 10% from the minimum cost, while the yellow circle marks the point where  $\kappa$  attains the maximum deviation.

The offline rollout and myopic+ policy parameters are shown in Fig. 4 for various values of  $\text{SNR}^2$ ,  $T = 20$ , and model parameters given by Table I. It should be noted that the offline rollout policies require numerical optimization over the  $\kappa(t)$  parameters, which tend to be noisy unless a large number of Monte Carlo realizations are used. In contrast, experiments in Section V indicate that the myopic+ policy parameters tend to be less sensitive to noise and mismodeling errors.

<sup>2</sup>SNR is defined in terms of the budget per stage  $\Lambda(t)$  and the noise variance  $\sigma^2$  as  $\text{SNR}(\Lambda(t)) = 10 \log_{10}(\Lambda(t)/(Q\sigma^2))$ .

```

procedure  $\{\kappa(\tau)\}_{\tau=1}^T = \text{Myopic+Policy}(\rho)$ 
  Set  $\kappa(1) = 1, \tilde{\kappa}(1) = \{\kappa(1)\}$ .
  for  $\tau = 2, 3, \dots, T - 1$  do
    for each  $\kappa(\tau) \in (0, 1]$  do
      Calculate  $B_\tau^D(\tilde{\kappa}(\tau - 1), \kappa(\tau))$  according to (25).
    end for
    Choose  $\hat{\kappa}(\tau)$  according to (24).
    Set  $\tilde{\kappa}(\tau) = \{\tilde{\kappa}(\tau - 1), \hat{\kappa}(\tau)\}$ .
  end for
  Return  $\{\kappa(\tau) = \{\tilde{\kappa}(\tau - 1), 0\}\}_{\tau=2}^T, \kappa(1) = 1$ .
end procedure

```

Fig. 2. Myopic+ policy pseudocode for determining exploration parameters  $\kappa$

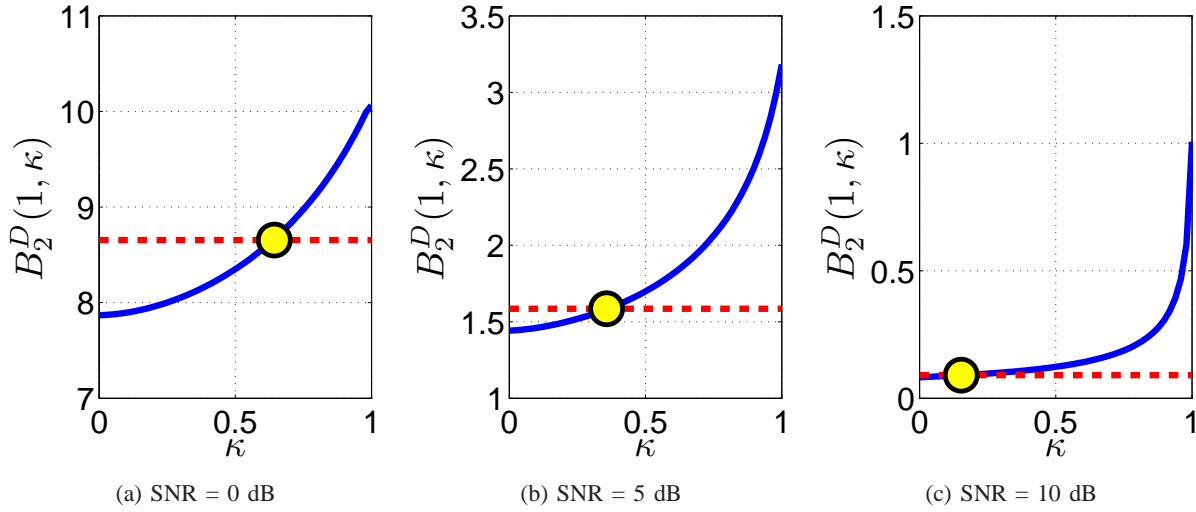


Fig. 3. We plot the myopic cost  $B_2^D(1, \kappa)$  as given by equation (25) as a function of the exploitive vs. explorative parameter  $\kappa$ .  $B_2^D(1, \kappa)$  is shown for low, medium, and high values of  $\Lambda(2)$  in (a), (b), and (c), respectively. In all cases, the myopic cost is optimized when  $\kappa = 0$ . However, lower SNR values can tolerate a larger value of  $\kappa$  and only have a small deviation in cost. The red dotted line shows a deviation of 10% from the minimum cost, while the yellow circle marks the point where  $\kappa$  attains this value.

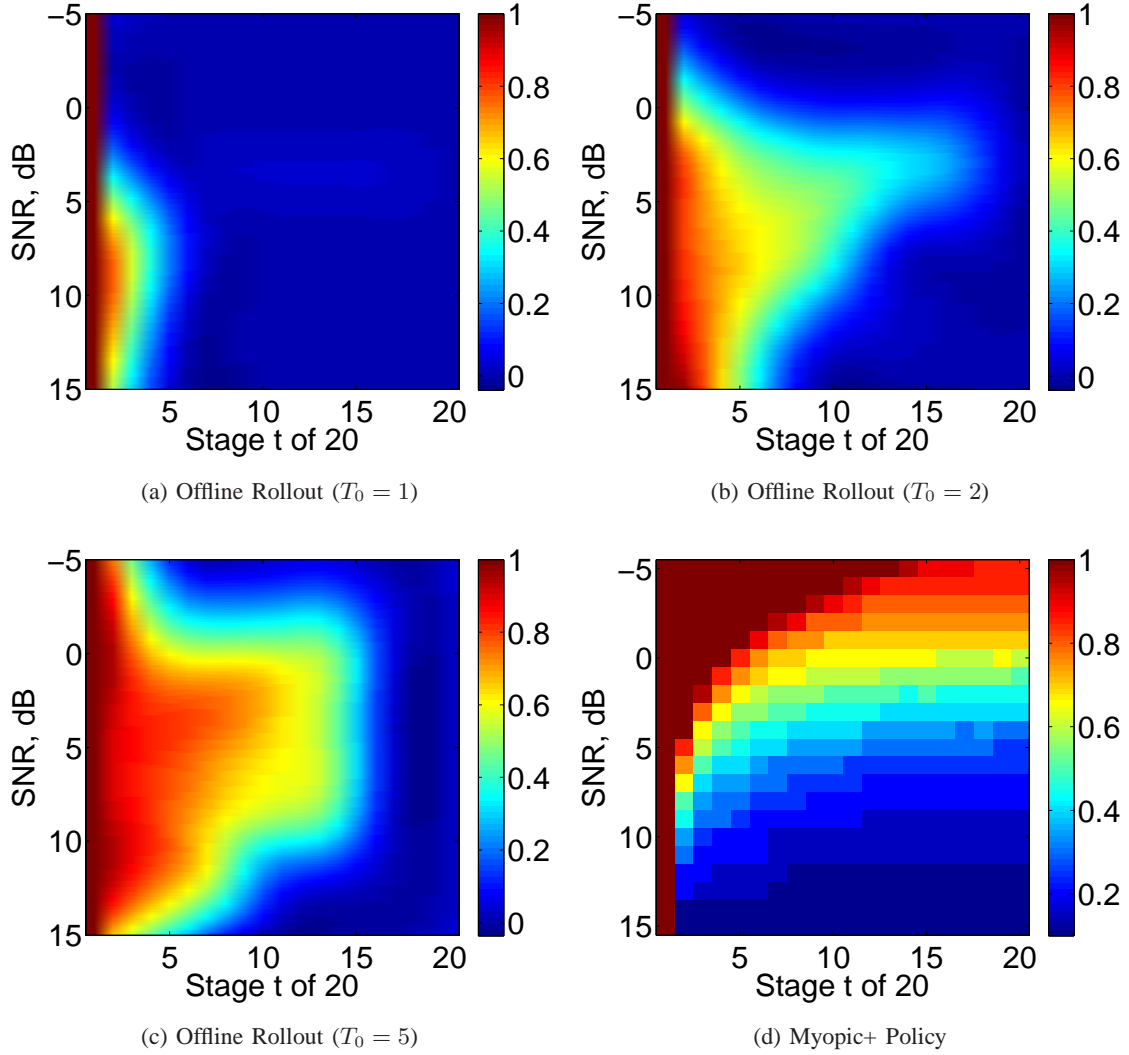


Fig. 4. The heat maps show the selection of the exploration coefficient  $\kappa(T)$  according to Algorithms 1 (offline rollout) and 2 (myopic+) for policies of length  $T = 20$ . The rollout policies use a myopic base policy (i.e.  $\kappa(t) = 0$ ) of lengths  $T_0 = 1, 2, 5$  in (a), (b), and (c), respectively. The myopic+ policy is given in (d), which performs more exploration in general for lower SNR than the rollout policies. In all four policies,  $\kappa(t)$  is nearly monotonically decreasing in  $t$ . The offline rollout policies all exhibit a phase transition from a low SNR regime (where  $\kappa(t) \approx 0$ ) to a higher SNR regime where significant exploration occurs. The heatmaps indicate that higher values of  $T_0$  lead to more exploratory strategies at low SNR and  $t$ . The myopic+ policy is monotonically decreasing in both SNR and  $t$ .

TABLE I  
PARAMETERS USED FOR SIMULATION ANALYSIS

Parameter	Variable Name	Value
Number of locations	$Q$	1,000
Prior sparsity	$p_i(1) = p_0$	0.01
Target amplitude mean	$\mu_i(1) = \mu_0$	1
Target amplitude std. deviation (prior)	$\sigma_i(1) = \sigma_0$	1/6
Target amplitude std. deviation (update)	$\Delta$	1/20
Noise variance	$\sigma^2$	1
Stationary probability	$\pi_0$	1/3
Death probability	$\alpha$	0
Birth probability	$\beta$	0
Number of neighbors	$ G $	2
Stage weights	$\{\gamma(t)\}_{t=1}^T$	$\{0, \dots, 0, 1\}$

#### IV. PERFORMANCE BOUNDS

In this section, we develop bounds on the performance gain that can be achieved with D-ARAP, and more generally any adaptive policy, compared to non-adaptive uniform allocation policies. The gain is measured using the cost function (10). The bounds result from analyzing two oracle policies that have exact knowledge of target locations. The first of these, the omniscient policy, has access to the target locations  $\Psi(t)$  for all  $t$  and is discussed in Sections IV-A and IV-C. The second, the semi-omniscient policy, has access to only the previous locations  $\Psi(t-1)$  at stage  $t$  and is considered in Sections IV-B and IV-D.

We distinguish two qualitatively different cases corresponding to either constant or increasing target amplitude variance, characterized by the increment  $\Delta^2 = 0$  or  $\Delta^2 > 0$  respectively. For oracle policies, the definitions of the state variables (6)–(8) are modified by augmenting the observation history  $\mathbf{Y}(t-1)$  with the exact target positions  $\Psi(t-1)$ , i.e.,  $\mathbf{Y}(t-1) \rightarrow \{\mathbf{Y}(t-1), \Psi(t-1)\}$ . In this case, it can be

shown that the posterior variances evolve according to

$$\sigma_i^2(t+1) = \frac{\sigma^2}{\sigma^2 + \lambda_{s^{(n)}(t)}(t)\sigma_{s^{(n)}(t)}^2(t)}\sigma_{s^{(n)}(t)}^2(t) + \Delta^2, \quad i \in H\left(s^{(n)}(t)\right), \quad (26)$$

where  $H(j) = \{j\} \cup G(j)$ . Hence in the case of static target amplitudes ( $\Delta^2 = 0$ , Sections IV-A and IV-B), the posterior variances decay to zero as  $t$  increases, while for  $\Delta^2 > 0$  (Sections IV-C and IV-D), the posterior variances reach a nonzero steady state. For simplicity, we make the following assumption for derivation of the performance bounds:

**Assumption 1.** The number of targets  $|\Psi(t)|$  is constant, i.e.,  $\alpha = \beta = 0$ .

*A. Omniscient policy,  $\Delta^2 = 0$*

In Sections IV-A and IV-B we make the additional assumption that the target amplitudes are constant:

**Assumption 2.** The variance increment  $\Delta^2$  is zero.

In this case, (26) reduces to a simple recursion for the posterior precisions  $c_i(t) = \sigma^2/\sigma_i^2(t)$ :

$$c_i(t+1) = c_{s^{(n)}(t)}(t) + \lambda_{s^{(n)}(t)}(t), \quad i \in H\left(s^{(n)}(t)\right), \quad (27)$$

where  $c_i(1) = \sigma^2/\sigma_0^2$  for all  $i$ .

The omniscient policy has perfect knowledge of the target locations  $\Psi(t)$  at all times. Conditioned on  $\Psi(t)$ , it follows that the target probabilities are atomic,  $p_i(t) = I_i(t)$ , and the omniscient policy allocates effort solely and uniformly to targets:

$$\lambda_i^o(t) = \begin{cases} \Lambda(t)/|\Psi(1)|, & i \in \Psi(t) \\ 0, & i \notin \Psi(t), \end{cases} \quad (28)$$

noting that  $|\Psi(t)| = |\Psi(1)|$  under Assumption 1. Given (27) and (28), the posterior precisions also remain uniform over targets:

$$c_i(t) = \frac{\sigma^2}{\sigma_0^2} + \frac{\bar{\Lambda}(t-1)}{|\Psi(1)|}, \quad i \in \Psi(t) \quad \forall t, \quad (29)$$

where  $\bar{\Lambda}(t) = \sum_{\tau=1}^t \Lambda(\tau)$ . To verify (28), we begin with  $t = 1$ , in which case  $c_i(1)$  is uniform over  $i \in \Psi(1)$ . Specializing the optimal allocation given by (18)–(20) to the case  $p_i(1) = I_i(1)$ , it is seen that the sequence  $g(k)$  (19) is equal to 0 for  $k < |\Psi(1)|$  and  $\infty$  for  $k \geq |\Psi(1)|$ . Hence the number of nonzero allocations  $k^* = |\Psi(1)|$  and (28) follows from (20). For  $t > 1$ , (28) continues to hold by induction since  $c_i(t)$  remains uniform over  $i \in \Psi(t)$ .

We define the gain of a policy with respect to the uniform allocation policy as

$$\Gamma_T(\boldsymbol{\lambda}) = \frac{J_T(\boldsymbol{\lambda}^u)}{J_T(\boldsymbol{\lambda})}. \quad (30)$$

Using (28) and (29), the gain of the omniscient policy is characterized in Proposition 1. The following assumption is used to obtain a more interpretable expression.

**Assumption 3.** The stage weights  $\gamma(t)$  decay to zero as  $t$  decreases from  $T$ .

This assumption ensures that as  $T \rightarrow \infty$ , the cost (10) becomes dominated by terms at large  $t$ . The assumption is satisfied by common “forgetting” schemes that emphasize performance in later stages.

**Proposition 1.** Let  $r_0(t) = \sigma^2 Q / (\sigma_0^2 \bar{\Lambda}(t))$ . Under Assumptions 1 and 2, the gain of the omniscient policy relative to uniform allocation is bounded from above as

$$\Gamma_T(\boldsymbol{\lambda}^o) \leq \left( \sum_{t=1}^T \frac{\gamma(t)}{\bar{\Lambda}(t)} \frac{1}{1 + r_0(t)} \right) / \left( \sum_{t=1}^T \frac{\gamma(t)}{\bar{\Lambda}(t)} \left[ \frac{p_0}{1 + p_0 r_0(t)} + \frac{1 - p_0}{Q} \frac{1}{(1 + p_0 r_0(t))^3} - \frac{(1 - p_0)(1 - 2p_0)}{Q^2} \frac{r_0(t)}{(1 + p_0 r_0(t))^4} \right] \right).$$

In the high-SNR limit ( $\sigma^2 \rightarrow 0$ ) or if Assumption 3 holds and the number of stages  $T \rightarrow \infty$ , then  $r_0(t) \rightarrow 0$  and the above expression simplifies to

$$\Gamma_T(\boldsymbol{\lambda}^o) = \frac{1}{p_0 + (1 - p_0)/Q} (1 - O(r_0)).$$

*Proof:* See Appendix II. ■

Proposition 1 shows that the omniscient gain is proportional to the sparsity of the scene, similar to [4], [5], [9]. In other words, the potential gain due to adaptation is higher when there are fewer targets.

### B. Semi-omniscient policy, $\Delta^2 = 0$

We now turn to the semi-omniscient policy, which in stage  $t$  has knowledge only of the previous target locations  $\Psi(t - 1)$ . In the semi-omniscient case, the target probabilities  $p_i(t) = \Pr(I_i(t) = 1 \mid \Psi(t - 1))$  are no longer binary but are given by the target dynamics (1) as

$$p_i(t) = \begin{cases} \pi_0, & i \in \Psi(t - 1), \\ \frac{1 - \pi_0}{|G|}, & i \in G \setminus \Psi(t - 1), \\ 0 & \text{otherwise,} \end{cases} \quad (31)$$



where  $G(\Psi(t-1)) = \bigcup_{i \in \Psi(t-1)} G(i)$  is the set of neighbors of all targets. We assume that the probability of target transitions is bounded.

**Assumption 4.** The probability of a target remaining in the same location is no smaller than the probability of it transitioning to any one neighboring cell,

$$\pi_0 \geq \frac{1 - \pi_0}{|G|}.$$

Unlike in the omniscient case, under the semi-omniscient policy the posterior precisions  $c_i(t)$  become random and non-uniform for  $t > 1$  over the set of locations  $H(\Psi(t-1)) = \Psi(t-1) \cup G(\Psi(t-1))$  where  $p_i(t) > 0$ . The non-uniformity arises because  $H(\Psi(t-1))$  contains both target and non-target locations, and even among targets, the precisions differ randomly depending on the number of times a target has stayed in the same cell or moved to a different one. This makes it difficult to determine the allocations analytically via (18)–(20). As an alternative, we focus on developing an upper bound  $\bar{c}(t)$  on the expected precisions  $\mathbb{E}[c_i(t) \mid |\Psi(1)|]$ ,  $i \in H(\Psi(t-1))$ , conditioned on the number of targets  $|\Psi(1)|$ . For  $t = 1$ ,  $\bar{c}(t)$  is defined as  $\bar{c}(1) = c_i(1) = \sigma^2/\sigma_0^2$ , satisfying the upper bound property. For  $t > 1$ ,  $\bar{c}(t)$  is defined by the recursion

$$\bar{c}(t+1) = \begin{cases} \frac{\pi_0^{3/2} + \frac{1}{\sqrt{|G|}}(1-\pi_0)^{3/2}}{\sqrt{\pi_0} + \sqrt{|G|(1-\pi_0)}} \left( (1 + |G|) \bar{c}(t) + \frac{\Lambda}{|\Psi(1)|} \right), & \bar{c}(t) < c_{\text{crit}}, \\ \bar{c}(t) + \frac{\pi_0 \Lambda}{|\Psi(1)|}, & \bar{c}(t) \geq c_{\text{crit}}, \end{cases} \quad (32)$$

where the threshold  $c_{\text{crit}}$  is defined as

$$c_{\text{crit}} = \frac{\Lambda}{|\Psi(1)|e(\pi_0, G)}, \quad (33)$$

and  $e(\pi_0, G) = \sqrt{\frac{|G|\pi_0}{1-\pi_0}} - 1 \geq 0$ .

We also use the following assumptions to determine the number of nonzero allocations under the semi-omniscient policy:

**Assumption 5.** The posterior precisions are uniform in the vicinity of targets,

$$c_i(t) = \bar{c}(t), \quad i \in H(\Psi(t-1)), \quad t > 1.$$

**Assumption 6.** The per-stage effort budget  $\Lambda(t)$  is constant,  $\Lambda(t) = \Lambda$ .

Assumption 5 replaces  $c_i(t)$  with an upper bound on its expected value and is therefore an optimistic approximation consistent with deriving the upper bound  $\bar{c}(t)$ . As  $t$  increases, the short-term deviations of  $c_i(t)$  from its mean decrease relative to the long-term increase of the mean and the approximation

corresponds to an upper bound on  $c_i(t)$  itself with high probability. We note that Assumption 5 is used primarily to determine the number of nonzero allocations and only indirectly to determine the amount allocated.

Given Assumptions 4–6, the following lemma proves that the recursion in (32) yields a valid upper bound on  $\mathbb{E}[c_i(t) \mid |\Psi(1)|]$ .

**Lemma 1.** *Under Assumptions 1, 2, and 4–6, the expected posterior precisions for the semi-omniscient policy satisfy*

$$\mathbb{E}[c_i(t) \mid |\Psi(1)|] \leq \bar{c}(t), \quad i \in H(\Psi(t-1)), \quad t \geq 1,$$

where  $\bar{c}(t)$  is defined by the recursion (32).

*Proof:* See Appendix III. ■

*Remark.* It can be shown that for  $\bar{c}(t) < c_{\text{crit}}$ , the coefficient multiplying  $\bar{c}(t)$  in (32) is greater than or equal to 1, with equality if and only if Assumption 4 holds with equality. Hence  $\bar{c}(t)$  increases geometrically with  $t$  if the inequality in Assumption 4 is strict. A closed-form expression can be derived for  $\bar{c}(t)$  in the regime  $\bar{c}(t) < c_{\text{crit}}$ , for example by viewing (32) as specifying a first-order recursive system driven by a step input, but we do not pursue this here.

Using Lemma 1 and taking the limit  $t \rightarrow \infty$ , we arrive at a simple characterization of the semi-omniscient policy.

**Proposition 2.** *Under Assumptions 1, 2, and 4–6, in the limit  $t \rightarrow \infty$  the expected per-stage cost of the semi-omniscient policy is bounded as*

$$\mathbb{E}[M_t(\boldsymbol{\lambda}^s)] \geq \frac{p_0 Q(p_0 Q + 1 - p_0)}{\pi_0 \Lambda t} + O\left(\frac{1}{t^2}\right).$$

*Proof:* See Appendix IV. ■

Proposition 2 can be used to determine the gain of the semi-omniscient policy relative to uniform allocation in the limit  $T \rightarrow \infty$ , again invoking Assumption 3 so that the total cost  $J_T(\boldsymbol{\lambda})$  is dominated by terms at large  $t$ . In the special case  $\gamma(T) = 1$ ,  $\gamma(t) = 0$  for  $t < T$ , the gain reduces to the ratio of the expected final-stage costs. From Proposition 2 and using (98) for the per-stage cost of the uniform policy with  $\bar{\Lambda}(t) = \Lambda t$ , the gain is bounded as

$$\Gamma_T(\boldsymbol{\lambda}^s) \leq \frac{\pi_0}{p_0 + (1 - p_0)/Q} + O\left(\frac{1}{T}\right) \approx \frac{\pi_0}{p_0}. \quad (34)$$

Compared to Proposition 1 in the limit  $T \rightarrow \infty$ , the analogous result for the omniscient policy, (34) shows that the performance of the semi-omniscient policy is discounted by the probability  $\pi_0$  that target locations are constant from stage to stage.

### C. Omniscient policy, $\Delta^2 > 0$

In the remainder of this section, we relax Assumption 2 on the variance increment  $\Delta^2$ . For  $\Delta^2 > 0$ , the evolution equation for posterior variances reverts to (26), from which it is difficult to obtain a closed-form expression for  $\sigma_{s(n)(t)}^2(t)$ , in contrast to the case  $\Delta^2 = 0$ . We focus instead on the steady-state behavior in the limit of large  $t$ . Using Assumption 3, in the limit  $T \rightarrow \infty$  the cost  $J_T(\boldsymbol{\lambda})$  becomes well-approximated by a sum of terms at large  $t$ , each of which is proportional to the steady-state expected per-stage cost  $\lim_{t \rightarrow \infty} \mathbb{E}[M_t(\boldsymbol{\lambda})]$ . This simplification allows us to obtain the following bound on the gain of the omniscient policy.

**Proposition 3.** *Let  $r_+ = \sigma^2 Q / (\Delta^2 \Lambda)$ . Under Assumptions 1, 3, and 6, in the steady-state limit  $T \rightarrow \infty$  the gain of the omniscient policy relative to uniform allocation is bounded from above as*

$$\begin{aligned} \lim_{T \rightarrow \infty} \Gamma_T(\boldsymbol{\lambda}^o) &\leq \left( \frac{\sqrt{1 + 4r_+} - 1}{2r_+} \right) / \\ &\quad \left( \frac{\sqrt{1 + 4p_0 r_+} - 1}{2r_+} + \frac{1 - p_0}{Q} \frac{1 + 3p_0 r_+}{(1 + 4p_0 r_+)^{3/2}} - \frac{(1 - p_0)(1 - 2p_0)}{Q^2} \frac{r_+(1 + 2p_0 r_+)}{(1 + 4p_0 r_+)^{5/2}} \right) \\ &= \frac{1}{p_0 + (1 - p_0)/Q} (1 - O(r_+)). \end{aligned}$$

*Proof:* See Appendix V. ■

Figure 5(a) is a heat map representing the bound in Proposition 3 as a function of SNR and  $p_0$ , where all other parameters are given in Table I. The upper horizontal axis indicates the equivalent values of  $r_+$ , which is inversely proportional to SNR when  $\Delta^2$  is fixed as in Table I. Besides confirming that the potential gain increases as  $p_0$  decreases, the heat map shows that there are three regimes with respect to SNR. In Region (A), the SNR is insufficient to offset the degradation due to  $\Delta^2$  and gains scale only as  $1/\sqrt{p_0}$ . In Region (C), the SNR is high and knowledge of target locations, which increases the observation effort per target by  $1/p_0$  on average, also increases the gain by approximately the same factor,  $1/[p_0 + (1 - p_0)/Q]$ . In Region (B), the gain ranges between the two extremes.

### D. Semi-omniscient policy, $\Delta^2 > 0$

Next we consider the steady-state behavior of the semi-omniscient policy. As discussed in Section IV-B, for  $t > 1$  the posterior precisions  $c_i(t)$  become non-uniform and random. However, in the regime

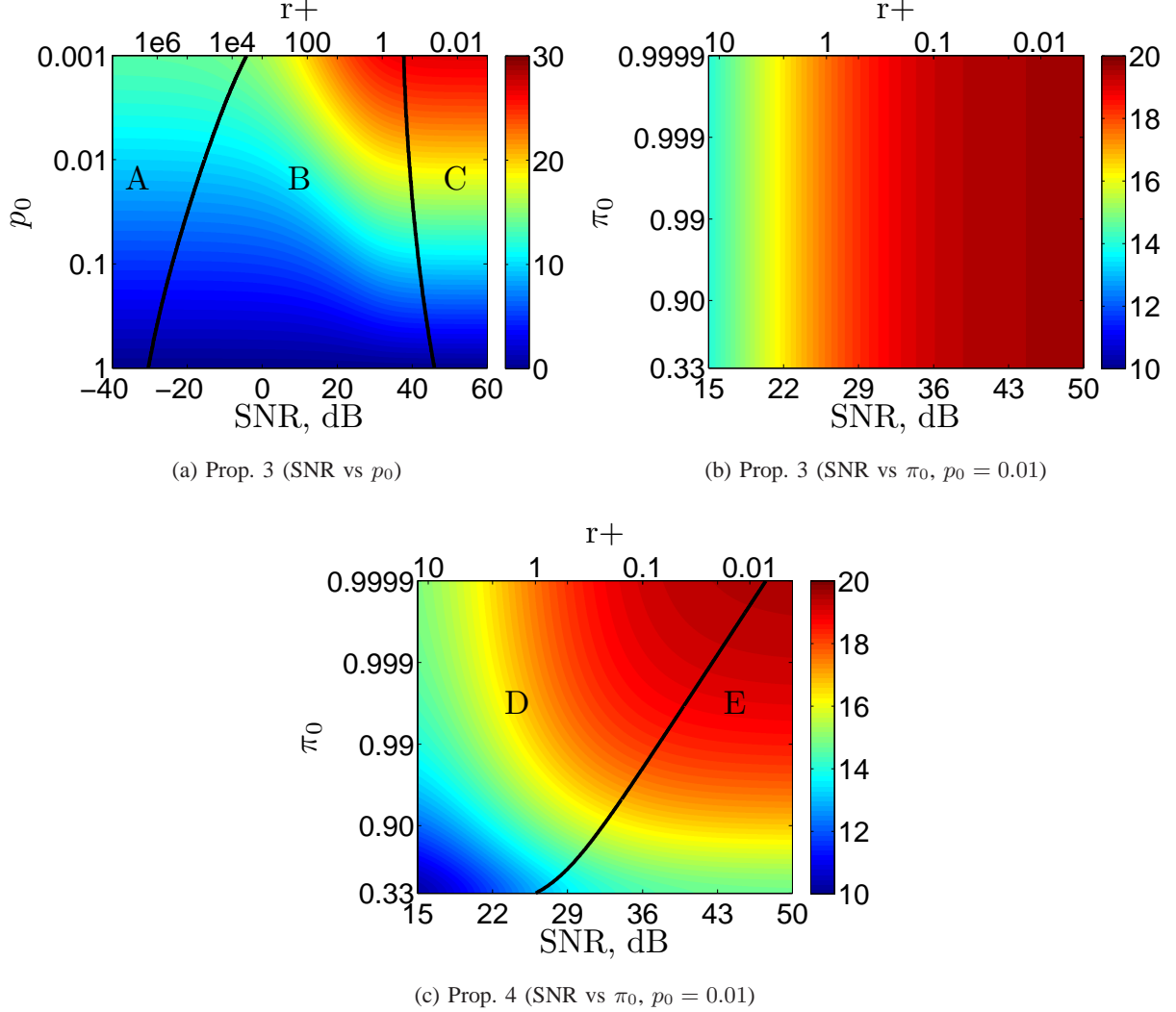


Fig. 5. Bounds on the gain of the omniscient policy (from Proposition 3, panels (a) and (b)) and the semi-omniscient policy (from Proposition 4, panel (c)) with respect to the uniform allocation policy. The left plot (a) confirms that potential gains increase as the sparsity parameter  $p_0$  decreases and indicates three regimes with respect to SNR: in Region (A), gains are relatively low at  $1/\sqrt{p_0}$ ; in Region (C), gains are near their maximum value  $1/[p_0 + (1 - p_0)/Q]$ ; in Region (B), they are in between. The middle and right plots (b, c) compare Props. 3 and 4 over the same values of SNR and  $\pi_0$ ; (E) denotes the region where the sufficient condition (35) is satisfied and Prop. 4 gives a provable upper bound. This bound (c) is tighter than the omniscient bound (b) because it accounts for the effect of having only causal knowledge of target locations, resulting in decreasing gains as  $\pi_0$  decreases.

of small  $r_+$ , where  $r_+$  is defined in Proposition 3, all  $c_i(t)$  are guaranteed to be small. We can then derive the following bound on the gain of the semi-omniscient policy under the same assumptions as in Proposition 3.

**Proposition 4.** Define  $q_+ = (1 + |G|)r_+$ . Given Assumptions 1, 3, and 6, assume in addition that

$$r_+(e(\pi_0, G) + 1) \leq 1. \quad (35)$$

Then in the steady-state limit  $T \rightarrow \infty$ , the gain of the semi-omniscient policy relative to uniform allocation is bounded from above as

$$\begin{aligned} \lim_{T \rightarrow \infty} \Gamma_T(\boldsymbol{\lambda}^s) &\leq \left( \frac{\sqrt{1 + 4r_+} - 1}{2r_+(\sqrt{\pi_0} + \sqrt{|G|(1 - \pi_0)})^2} \right) / \\ &\quad \left( \frac{p_0}{1 + p_0q_+} + \frac{1 - p_0}{Q} \frac{1}{(1 + p_0q_+)^3} - \frac{(1 - p_0)(1 - 2p_0)}{Q^2} \frac{q_+}{(1 + p_0q_+)^4} \right) \\ &= \frac{1 - O(r_+)}{(p_0 + (1 - p_0)/Q) \left( \sqrt{\pi_0} + \sqrt{|G|(1 - \pi_0)} \right)^2}, \end{aligned}$$

with equality in the limit  $r_+ \rightarrow 0$ .

*Proof:* See Appendix VI. ■

Figures 5(b) and (c) compare the omniscient and semi-omniscient bounds in Propositions 3 and 4 as functions of SNR and  $\pi_0$ . All other parameters are fixed as in Table I. The sufficient condition (35) for Proposition 4 is satisfied in region (E) in Fig. 5(c). The resulting semi-omniscient bound is tighter than the omniscient bound because it accounts for the lack of knowledge of future target locations, reflected in decreasing gains as  $\pi_0$  decreases. In the limit  $r_+ \rightarrow 0$ , a comparison of Propositions 3 and 4 shows that the semi-omniscient vs. omniscient degradation factor is  $\left( \sqrt{\pi_0} + \sqrt{|G|(1 - \pi_0)} \right)^{-2}$ , which is strictly less than 1 for  $\pi_0 < 1$ . Comparisons of Propositions 3 and 4 to the proposed D-ARAP policies are presented in Section V.

For the general case  $\Delta^2 > 0$ , we take a similar approach as in Section IV-B, invoking Assumptions 4–6 to obtain an upper bound  $\bar{c}(t)$  on the conditional expected posterior precisions  $\mathbb{E}[c_i(t) \mid |\Psi(1)|]$ . As before, there are two regimes to consider,  $\bar{c}(t) < c_{\text{crit}}$  and  $\bar{c}(t) \geq c_{\text{crit}}$ , where  $c_{\text{crit}}$  is given in (33). For simplicity, we restrict attention to the second regime and establish the following result to propagate the upper bound  $\bar{c}(t)$  forward in time.

**Lemma 2.** Under Assumptions 1 and 4–6, if the expected posterior precisions under the semi-omniscient

policy satisfy

$$\mathbb{E}[c_i(t) \mid |\Psi(1)|] \leq \bar{c}(t), \quad i \in H(\Psi(t-1)), \quad (36)$$

for some  $t$  with  $\bar{c}(t) \geq c_{\text{crit}}$ , then (36) also holds for stage  $t+1$  with

$$\bar{c}(t+1) = \pi_0 \frac{(\sigma^2/\Delta^2)(\bar{c}(t) + \Lambda/|\Psi(1)|)}{(\sigma^2/\Delta^2) + \bar{c}(t) + \Lambda/|\Psi(1)|} + (1 - \pi_0) \frac{(\sigma^2/\Delta^2)\bar{c}(t)}{(\sigma^2/\Delta^2) + \bar{c}(t)}. \quad (37)$$

*Proof:* See Appendix VII. ■

*Remark.* As  $\Delta^2 \rightarrow 0$ , the above recursion (37) reduces to (32) in the case  $\bar{c}(t) \geq c_{\text{crit}}$ . Hence (37) can be seen as a generalization of (32) to the case  $\Delta^2 > 0$ .

Lemma 2 can be used to derive an upper bound on the expected posterior precisions in the steady state limit  $t \rightarrow \infty$ . Define  $c_{\text{ss}} = \lim_{t \rightarrow \infty} \mathbb{E}[c_{s(n)(t)}(t) \mid |\Psi(1)|]$  to be the steady-state precision for targets.

**Lemma 3.** *Under Assumptions 1 and 4–6, the steady-state precision  $c_{\text{ss}}$  for targets is bounded from above by a root  $\bar{c}_{\text{ss}}$  of the cubic equation*

$$\pi_0 \bar{c}_{\text{ss}}^{-3} - \frac{\Delta^2}{\sigma^2} \left( \frac{\Delta^2}{\sigma^2} + \frac{|\Psi(1)|}{\Lambda} \right) \bar{c}_{\text{ss}}^{-1} - \left( \frac{\Delta^2}{\sigma^2} \right)^2 \frac{|\Psi(1)|}{\Lambda} = 0, \quad (38)$$

provided that  $\bar{c}_{\text{ss}} \geq c_{\text{crit}}$ .

*Proof:* Initially, let  $\bar{c}_{\text{ss}}$  denote an upper bound on  $c_{\text{ss}}$  such that  $\bar{c}_{\text{ss}} \geq c_{\text{crit}}$ . Then Lemma 2 implies that  $c_{\text{ss}}$  is also bounded by

$$\pi_0 \frac{(\sigma^2/\Delta^2)(\bar{c}_{\text{ss}} + \Lambda/|\Psi(1)|)}{(\sigma^2/\Delta^2) + \bar{c}_{\text{ss}} + \Lambda/|\Psi(1)|} + (1 - \pi_0) \frac{(\sigma^2/\Delta^2)\bar{c}_{\text{ss}}}{(\sigma^2/\Delta^2) + \bar{c}_{\text{ss}}}.$$

To obtain a stationary upper bound on  $c_{\text{ss}}$ , we equate  $\bar{c}_{\text{ss}}$  with the above expression, resulting in (38) after some algebraic manipulations. This stationary bound is valid provided that a root of (38) satisfies the initial assumption  $\bar{c}_{\text{ss}} \geq c_{\text{crit}}$ . ■

In general, it is difficult to obtain a tractable expression for  $\bar{c}_{\text{ss}}$  from (38). We consider two special cases. In the case  $\pi_0 = 1$ , the cubic polynomial can be factored into

$$\left( \bar{c}_{\text{ss}}^{-1} + \frac{\Delta^2}{\sigma^2} \right) \left( \bar{c}_{\text{ss}}^{-2} - \frac{\Delta^2}{\sigma^2} \bar{c}_{\text{ss}}^{-1} - \frac{\Delta^2}{\sigma^2} \frac{|\Psi(1)|}{\Lambda} \right).$$

The first factor yields an infeasible negative root while the second factor can be shown to be proportional to the quadratic polynomial in (105) with  $\lambda_{s(n)(t)}(t) = \Lambda/|\Psi(1)|$ , the effort allocation for targets under the omniscient policy. Thus setting  $\pi_0 = 1$  recovers the omniscient case analyzed in Section IV-C. For  $\pi_0 < 1$ , a tractable solution can also be extracted if  $\Delta^2$  is close to zero, as detailed in the following lemma.

**Lemma 4.** *In the limit  $\sqrt{\Delta^2\Lambda/\sigma^2} \rightarrow 0$ , the cubic equation (38) has a single positive root given by*

$$\bar{c}_{\text{ss}}^{-1} = \sqrt{\frac{\Delta^2|\Psi(1)|}{\pi_0\sigma^2\Lambda}} \left( 1 + O\left(\sqrt{\frac{\Delta^2\Lambda}{\sigma^2}}\right) \right). \quad (39)$$

*Proof:* See Appendix VIII. ■

Combining Lemmas 3 and 4, we arrive at the following steady-state characterization of the semi-omniscient policy.

**Proposition 5.** *Given Assumptions 1 and 3–6, assume in addition that  $\sqrt{\Delta^2\Lambda/\sigma^2} \ll 1$  and  $\pi_0$  is large enough so that*

$$\sqrt{\frac{\pi_0\sigma^2}{\Delta^2\Lambda}} e(\pi_0, G) \geq 1. \quad (40)$$

*Then in the steady-state limit  $T \rightarrow \infty$ , the gain of the semi-omniscient policy relative to uniform allocation is bounded from above as*

$$\lim_{T \rightarrow \infty} \Gamma_T(\boldsymbol{\lambda}^s) \leq \sqrt{\frac{\pi_0}{p_0}} \left( 1 + O\left(\sqrt{\frac{\Delta^2\Lambda}{\sigma^2}}\right) \right).$$

*Proof:* See Appendix IX. ■

We again compare the above result to Proposition 3 on the omniscient policy, this time in the regime  $r_+ \gg \sigma^2/(\Delta^2\Lambda) \gg 1$ . In this case, the gain of the omniscient policy is approximately given by  $1/\sqrt{p_0}$  and the performance loss due to causal knowledge of target motion is therefore  $\sqrt{\pi_0}$ .

## V. NUMERICAL SIMULATIONS

### A. Simulation set-up

In this section, we analyze the performance of the proposed rollout and myopic+ D-ARAP policies in a variety of situations that include model mismatch and missing measurements. We further examine the policies over a variety of performance metrics (MSE and probability of detection). With regard to rollout policies, we investigate the effects of using different base policies for offline rollout, and also compare the performance of offline and online rollout. We continue by investigating the sensitivity of the dynamical model by varying birth/death probabilities and transition probabilities. Simulation parameters are given by Table I unless stated otherwise.

### B. Comparison to semi-omniscient/uniform policies

In this section, we examine the performance of all of the proposed policies (offline rollout, myopic+, and myopic) as well as the semi-omniscient oracle, which provides an upper bound on performance.

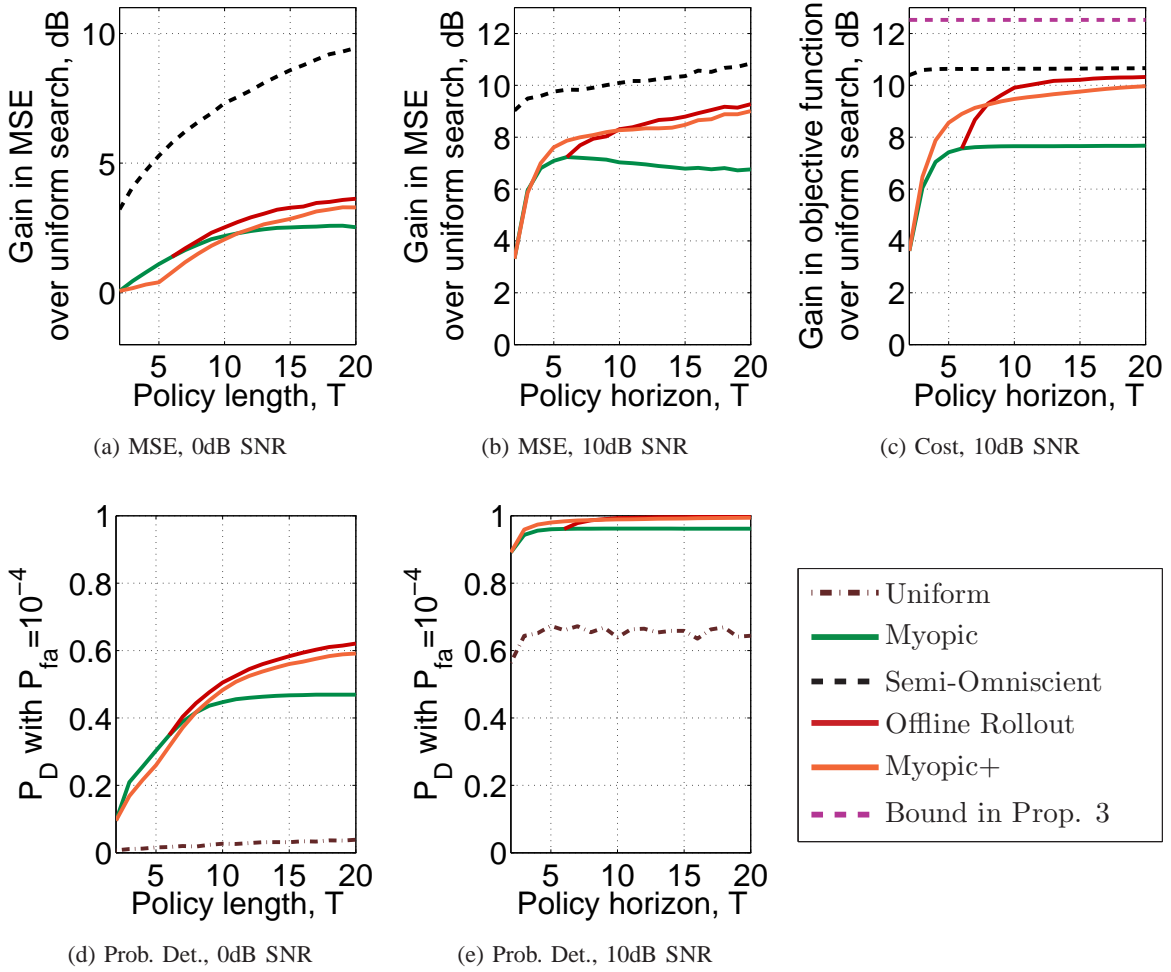


Fig. 6. These plots compare estimation and detection performance with 0 dB and 10 dB SNR for various values of  $T = 1, 2, \dots, 20$ . In (a) and (b), gains in MSE are plotted with respect to a uniform allocation policy (on a dB scale) for 4 alternative policies: the myopic policy (green), the myopic+ policy (orange), the offline rollout policy with  $T_0 = 5$  (red), the semi-omniscient oracle policy (black dashed), and the upper bound from Proposition 3 (purple dashed). In (c), the cost in (10) is plotted with respect to a uniform allocation policy. In (d) and (e), the detection probability with fixed false alarm rate of  $10^{-4}$  is shown for the same policies as well as the uniform allocation policy (brown dash-dotted). Observe that the proposed myopic+ policy and the offline rollout policy perform best, but the myopic+ policy has significantly lower implementation complexity.

Fig. 6 (a) and (b) show the MSE gains (with respect to a uniform policy) for estimating  $\{\theta_i(T)\}_{i \in \Psi(T)}$  for different values of  $T$  with 0 dB and 10 dB SNR, respectively. Generally the offline rollout policy has the highest gains in MSE among non-oracle policies, with performance close to the semi-omniscient policy as  $T$  gets large. However, the performance gain of the offline rollout is small with respect to the myopic+ policy. Fig. 6(c) provides the gain in the objective function in (10) with respect to the uniform



policy. Recall that (10) is used as a surrogate optimization objective for amplitude estimation MSE. Comparing (b) and (c), it is clear that improvements in cost generally lead to improvements in MSE, suggesting that (10) is a good surrogate function. In (c), the bound in Prop. 3 is also plotted (note that the condition (35) in Prop. 4 is not satisfied). In the next section, we empirically analyze the conditions which lead to tight oracle bounds as a function of model parameters and SNR.

Figs. 6 (d) and (e) show the probability of detection for a fixed probability of false alarm ( $P_{fa} = 10^{-4}$ ) as a function of  $T$  for 0 dB and 10 dB SNR, respectively. The probability of detection for D-ARAP (offline rollout and myopic+ policies) consistently approaches 1 as  $T$  gets large and does so significantly faster than for the uniform and myopic policies. Moreover, D-ARAP achieves perfect detection  $P_d = 1$  within just a few stages.

### C. Comparison across dynamic model parameters

We continue by comparing the performance of the myopic, myopic+, and semi-omniscient policies as a function of the dynamic model parameters. Figs. 7(a) and (b) analyze performance as a function of the transition probability  $(1 - \pi_0)$  and the number of potential neighbors  $|G|$ . In (a), we fix SNR=20 dB and vary  $\pi_0$ . Solid lines represent the case where  $|G| = 2$  while dashed lines represent  $|G| = 4$ . Moreover, we compare to Propositions 3 and 4 in the regimes where they apply, and extrapolate in between by taking the minimum of the two bounds. In (b) we compare across the semi-omniscient, myopic+ and myopic policies when fixing  $\pi_0 = 1/3$  and varying SNR. In (a), the theoretical bounds are tight in comparison to both the numerical semi-omniscient policy and the adaptive policies. In (b), the theoretical bounds are tight when the sufficient SNR condition (35) of Prop. 4 is nearly satisfied. Also, we see that the myopic policy tends to have poorer performance than the myopic+ policy when either SNR is low or  $\pi_0$  is low.

Figs. 7(c) and (d) provide a comparison as a function of the target birth probability,  $\beta$ . In these simulations, we set the target death probability,  $\alpha$ , as a function of  $\beta$  to keep the expected number of targets the same across stages. If  $\beta$  increases, then more resources are allocated uniformly across the scene to search for new targets. This results in a reduction in gains compared to the uniform search. This pattern is verified in Fig. 7(c) for fixed SNR=20 dB. The semi-omniscient policy performs the best, followed by the myopic+ and myopic policies. We also provide a theoretical upper bound as given in Chapter 3 of [15], which was derived under stronger assumptions than given in Section IV. There is an apparent bias between the semi-omniscient policy and this theoretical upper bound. In (d), we analyze this bias as a function of SNR while fixing  $\beta = 0.01$ . As SNR gets large, all policies approach the theoretical upper bound on performance.

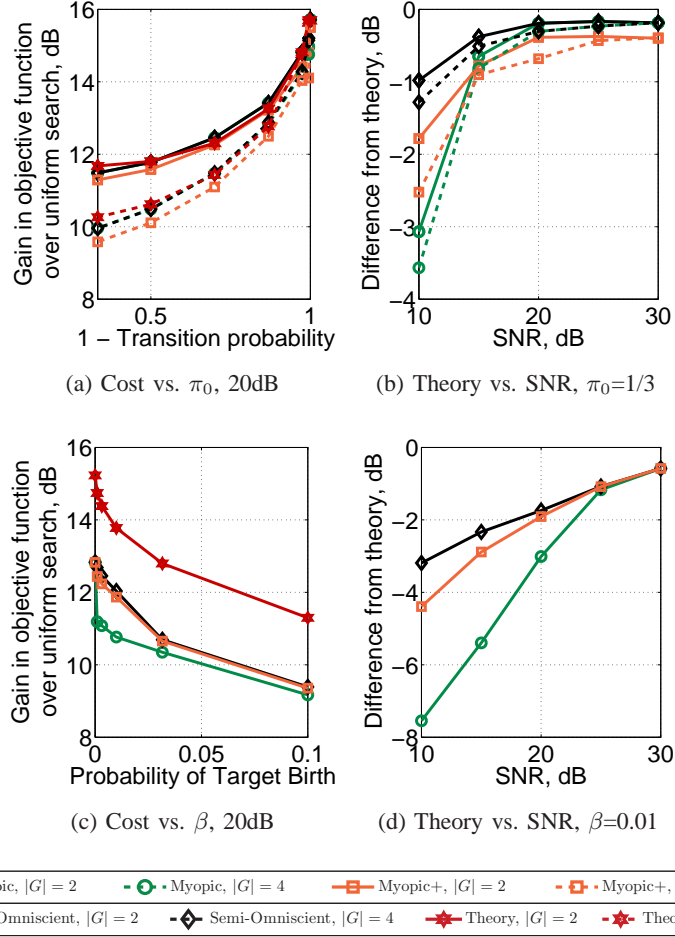


Fig. 7. These figures compare the performance of policies as a function of the dynamic model parameters. (a) and (b) show comparisons with variations in the transition probability ( $1 - \pi_0$ ) and the number of neighbors  $|G|$ . Solid lines indicate the performance when  $|G| = 2$ , while dashed lines indicate performance when  $|G| = 4$ . (b) compares to the minimum of the theoretical bounds in Propositions 3 and 4 when fixing  $\pi_0 = 1/3$  and varying SNR. (c) and (d) provide similar analysis as a function of the target birth probability,  $\beta$ . (c) shows results when fixing SNR=20 dB, while (d) compares the policies to the upper bound given in Chapter 3 of [15] as a function of SNR while fixing  $\beta = 0.01$ . In (a), the theoretical bounds are tight in comparison to both the numerical semi-omniscient policy and the adaptive policies. In (b), the theoretical bounds are tight when the sufficient SNR condition (35) of Prop. 4 is nearly satisfied. Similarly, in (d), the error between theoretical and numerical performance is also non-zero for small SNR, yet decreases near to zero as SNR increases.

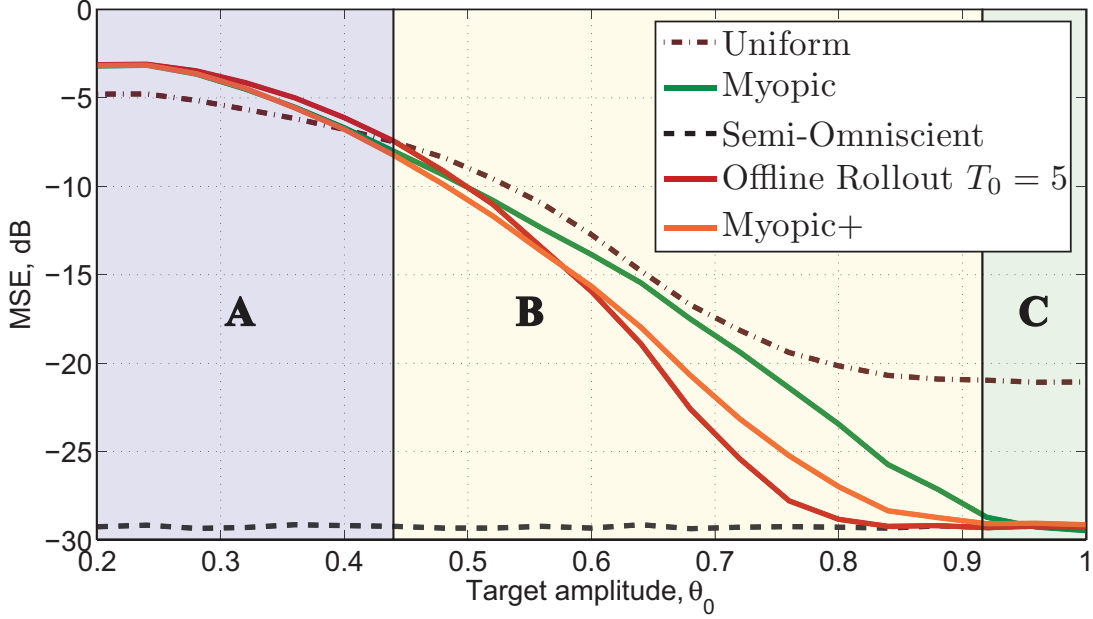


Fig. 8. This figure compares the proposed non-myopic policies (offline rollout and myopic+) to the myopic and semi-omniscient policies in a model mismatch scenario. Policies are derived under the assumption that the mean target amplitude  $\mu_0 = 1$ . However, these policies are mismatched to the actual target amplitudes  $\theta_i(t)$ , which are constant and lower than expected with  $\theta_i(t) = \theta_0 < 1$ . The figure plots MSE for various  $\theta_0$  (x-axis) and policies (curves). This figure is divided into 3 regions: in Region (A), there is not sufficient signal for adaptive policies to outperform the uniform alternative; in Region (B), adaptive policies perform better, yet there is significant benefit in using non-myopic strategies; in Region (C), all adaptive policies perform similarly.

#### D. Model Mismatch

In this section, we compare the non-myopic policies to the myopic and semi-omniscient oracle policy in cases where there might be model mismatch. In particular, we consider the case where the policy is derived under the model given in Table I with prior mean amplitude  $\mu_0 = 1$ . This model is mismatched to the actual measurements where the target amplitudes are all identical and lower than expected

$$\theta_i(t) = \theta_0 < \mu_0 = 1 \quad (41)$$

for  $\theta_0 \in [0.2, 1]$ . For  $\theta_0 \ll 1$ , noisy measurements from cells containing targets can be easily confused with the background noise. In these situations, the myopic policy will be more adversely affected by small posterior probability  $p_i(t)$  in the ROI  $i \in \Psi$  as compared to the non-myopic policies. Fig. 8 shows the performance as a function of MSE (lower is better). When  $\theta_0 < 0.45$  (Region A), all policies perform worse than the uniform search, indicating severe model mismatch. Conversely all policies show positive

performance for  $\theta_0 > 0.45$  (Regions B and C). Moreover, the performances of the non-myopic policies improve at a faster rate than the myopic policy, and quickly approach the theoretical bound as given by the semi-omniscient oracle policy.

#### *E. Complex dynamic behavior: missing measurements*

In the next simulation, we test the policies in the scenario where the sensor turns off periodically for several consecutive stages, creating time periods of missing data. This is representative of a modern radar system that must multi-task between different modes of operation, e.g., tracking, automated target recognition, and synthetic aperture radar [19], which compete for radar resources. In this simulation, 6 stages of data are collected followed by 3 stages of no measurements. Fig. 9 shows the resultant MSE (on a dB scale, lower is better) for the uniform, myopic, offline rollout, myopic+, and semi-omniscient policies and for various SNR (per-stage budgets of 0, 5, 10 and 15 dB).

All adaptive policies perform better than uniform search which has growing errors with  $t$ . In Fig. 9(a) and Fig. 9(b) (very low and low SNR), the non-myopic policies outperform the myopic and uniform policies, though their errors still grow with  $t$ . With sufficient SNR, in Fig. 9(c) (medium SNR) the non-myopic policies approach a stable point over each period of measurements, while the myopic policy progressively increases over those time points. For high SNR (15 dB) in (d), all adaptive policies perform similarly and close to the semi-omniscient oracle.

#### *F. Offline and online rollout policies*

Fig. 10 compares estimation and detection performance for the offline rollout policies as a function of the base policy. We consider myopic base policies which set the exploration parameter  $\kappa(t) = 0$  for  $T_0 = 1, 2, 5$  consecutive stages. We compare performance for two SNR levels, with SNR=10 dB given by diamonds, and SNR=0 dB given by circles. All policies perform at least as well as the myopic policy. Moreover, increasing  $T_0$  generally improves performance in both estimation and detection. It should be noted that for low SNR, using  $T_0 = 1$  did not noticeably improve performance over the myopic policy.

As discussed in the introduction, the full POMDP solution to the adaptive sensing problem is generally intractable due to the size of belief state and action spaces. As an alternative, we consider an approximate POMDP solution, namely the (online) rollout policy, in order to compare D-ARAP to online solutions. Note that in online solutions, the optimal action at each stage must be chosen separately for each realization of the model. Thus, the online rollout policy likely will incur significant computational costs in comparison to the offline policies presented in this paper.

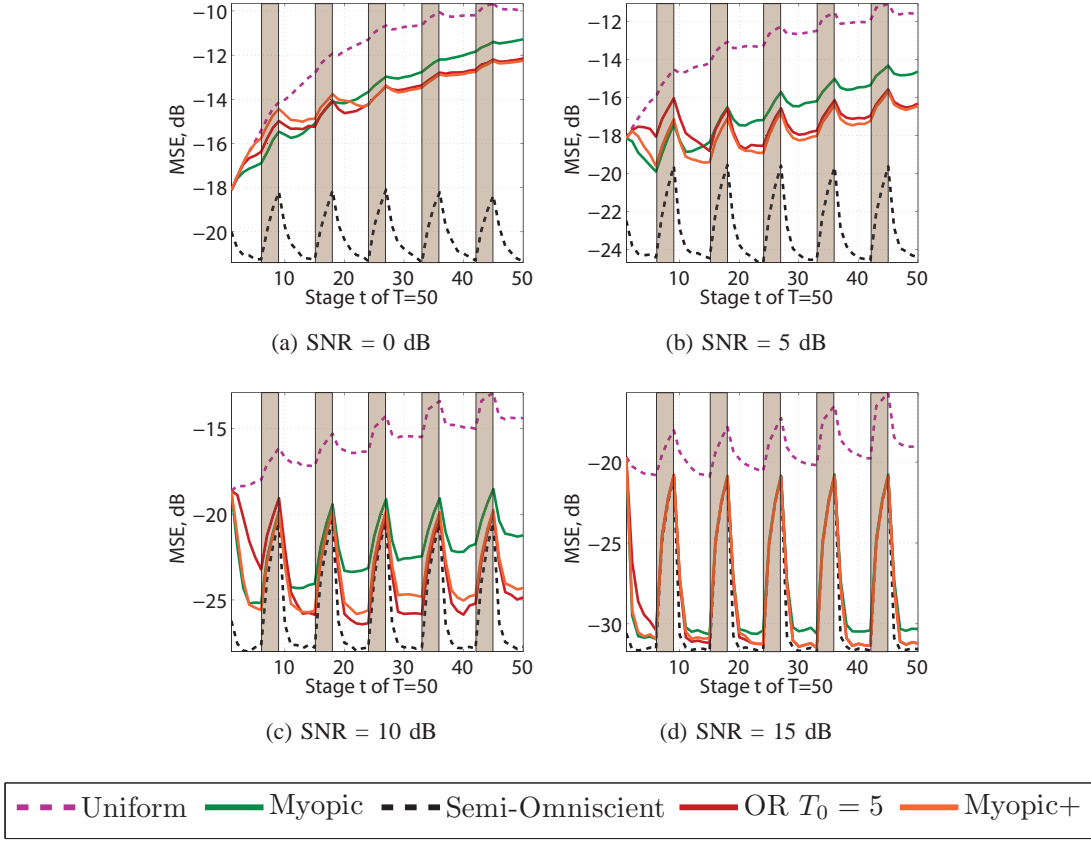


Fig. 9. This figure illustrates the effect of missing data on MSE performance of the allocation policies. In this scenario, 6 stages of observations are followed by no measurements for three consecutive stages (shaded regions). We report MSE on a dB scale (lower is better) for non-myopic, myopic, semi-omniscient, and uniform policies. Performance is compared for four SNR levels in 0, 5, 10, and 15 dB. Note that adaptive policies enable MSE to stabilize in the absence of measurements when there is sufficient SNR as in (c) and (d), whereas the uniform search has growing errors.

We compare offline and online policies using myopic base policies of various stage lengths,  $T_0$ . In Fig. 11(a), we compare the performance of the offline and online rollout policies in the SNR=10dB case for parameters given in Table I. The online and offline rollout policies perform similarly in the standard model (a) without model mismatch or missed observations. The online policy has significantly noisier results, which is partly caused by computational limits on the number of realizations from which the average performance is computed. Nevertheless, the online policy clearly performs better than the myopic policy. In Fig. 11(b), we compare the performance of the offline and online rollout policies in the SNR=10 dB case where stages of measurements are missing as in Section V-E. It is seen that the online  $T_0 = 2$  rollout policy performs similarly to the offline  $T_0 = 5$  policy. On the other hand, the

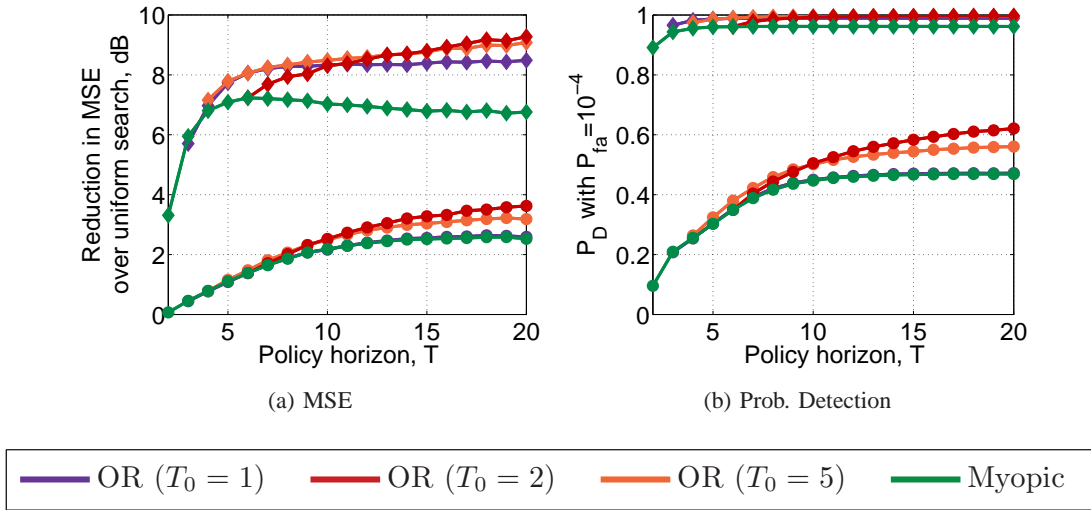


Fig. 10. These plots compare estimation and detection performance for the offline rollout policies as a function of the base policy. We consider myopic policies which set  $\kappa(t) = 0$  for  $T_0 = 1, 2, 5$  consecutive stages. Performance is shown for two SNR levels, with SNR=10 dB given by diamonds, and SNR=0 dB given by circles. It is seen that higher values of  $T_0$  (i.e., less myopic policies) tend to perform better in terms of both estimation and detection error.

online  $T_0 = 5$  policy performs significantly worse (and approximately the same as the myopic policy).

In our experience, the online rollout policies tend to be significantly noisier than their offline counterparts. This may be due to (a) necessary tradeoffs in computational (Monte Carlo) effort vs. accuracy or (b) difficulties in sampling from the belief state, particularly in sparse scenarios where the probability of targets existing at given locations is small. This indicates one advantage of the offline policies, which tend to be more robust to complex environments.

## VI. DISCUSSION AND FUTURE WORK

This paper presented a framework for adaptive sampling that significantly extends previous work [4] to incorporate dynamic targets while providing a computationally tractable solution, namely D-ARAP. A cost function related to mean squared error was proposed and upper bounds on the performance of adaptive sensing were derived through analysis of oracle policies. These bounds shed light on the impact of target motion and amplitude variation in addition to target sparsity. In terms of implementable policies, a myopic solution is given that has an analytical form, but suffers from being overly aggressive in the allocation of resources in cases of model mismatch or faulty measurements. We offer a non-myopic extension of the myopic policy that balances exploration of the scene and exploitation of prior

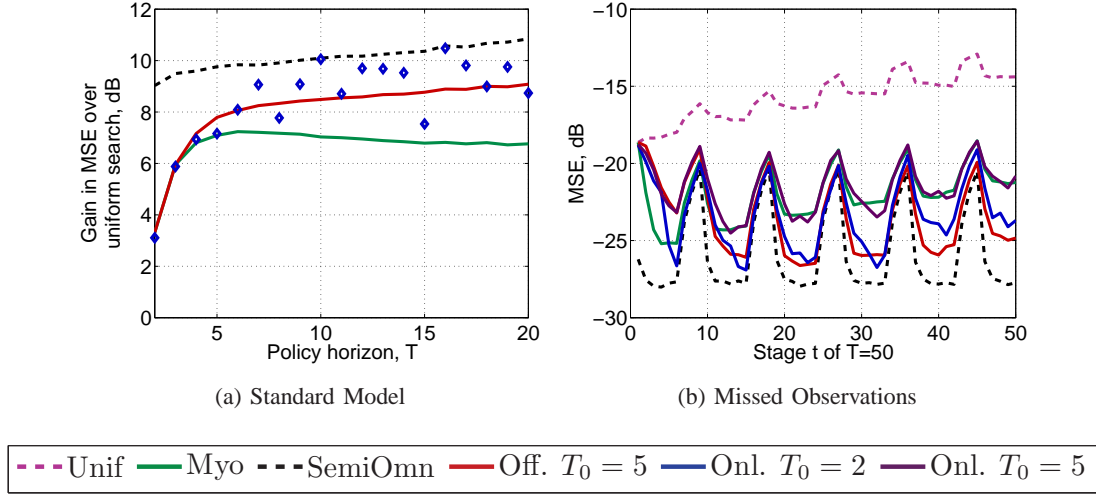


Fig. 11. These plots compare offline and online rollout policies in two scenarios. In (a), we compare the performance of the offline and online rollout policies in the SNR=10dB case for parameters given in Table I. In (b), we compare the performance of the offline and online rollout policies in the SNR=10 dB case where stages of measurements are missing as in Section V-E. In the standard model (a), the online policy performs similarly (albeit with more noise) to the offline version. In (b), however, the online policies performs worse than the offline version, most likely due to the additional computational complexity of the online policy. Nevertheless, the online policy with  $T_0 = 2$  still performs better than the myopic policy.

observations in a tractable manner. Numerical evidence suggests that the proposed D-ARAP policies (a) have significant performance gains over the baseline policy that uniformly allocates resources across the scene, (b) perform similarly to the gold-standard POMDP approximate solutions, albeit at a fraction of the computational cost, and (c) improve upon the myopic policy, especially in terms of robustness to model mismatch and faulty measurements.

Future research directions include consideration of constraints on the number of measurements, which may include coarse-scale or compressed sensing measurements. Moreover, further analytical results are of interest, for example convergence rates (in comparison to exhaustive search) and/or minimum detectable amplitudes, and performance bounds that are more refined than the oracle bounds presented herein. Finally, online policies that are computed as measurements are taken may be worthy of continued study since they could improve performance dramatically in some cases, including cases where targets will be obscured in the near future.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Eran Bashan for his valuable insights and assistance in the development of this research.

## APPENDIX I

## EFFICIENT POSTERIOR ESTIMATION FOR GIVEN DYNAMIC STATE MODEL

In order to use the algorithms provided in this work to adaptively estimate the state  $\xi(t)$  given the measurements, we need to be able to calculate the posterior probabilities for the indicator variables,  $\{I_i(t)\}_{i=1}^Q$  given the measurements up until time  $t$ . To do this efficiently (which is required for planning purposes), we make the following assumptions:

**Assumption 7.** There is at most 1 target in the vicinity of any target:

$$(s^{(n)} = i) \rightarrow |G(i)| = 0, \quad \forall n$$

where  $G(i)$  is the set of neighbors of location  $i$ .

Define the measurement vectors

$$\mathbf{y}(t) = \{y_1(t), y_2(t), \dots, y_Q(t)\} \quad (42)$$

and

$$\mathbf{Y}(t) = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(t)\} \quad (43)$$

Let

$$\Pr(I_i(t) = 1 | \mathbf{Y}(t-1)) \quad i = 1, 2, \dots, Q, \quad t = 1, 2, \dots, T \quad (44)$$

be the posterior probabilities that need to be calculated. For  $t = 1$ , we have

$$\Pr(I_i(1) = 1) = p, \quad (45)$$

under assumption 7. For  $t > 1$ , we have

$$\begin{aligned} \Pr(I_i(t) = 1 | \mathbf{Y}(t-1)) &= \int \Pr(I_i(t) = 1, \mathbf{S}(t) | \mathbf{Y}(t-1)) d\mathbf{S}(t) \\ &= \int \Pr(I_i(t) = 1 | \mathbf{S}(t)) f(\mathbf{S}(t) | \mathbf{Y}(t-1)) d\mathbf{S}(t) \\ &= \sum_{n=1}^{N(t)} \Pr(s^{(n)}(t) = i | \mathbf{Y}(t-1)), \end{aligned} \quad (46)$$



where the last equation can be derived noting that

$$\Pr(I_i(t) = 1 | \mathbf{S}(t)) = \begin{cases} 1, & \exists n : s^{(n)}(t) = i \\ 0, & \text{else} \end{cases} \quad (47)$$

Thus, in order to compute equation (46), we need to be able to estimate the state  $\xi(t)$  given  $\mathbf{Y}(t-1)$ .

#### A. Recursive equations for updating $\xi(t)$

In general, we can compute the posteriors using the equations:

$$f(\xi(t) | \mathbf{Y}(t-1)) = \int f(\xi(t) | \xi(t-1)) f(\xi(t-1) | \mathbf{Y}(t-1)) d\xi(t-1) \quad (48)$$

$$f(\xi(t) | \mathbf{Y}(t)) = \frac{f(\mathbf{y}(t) | \xi(t)) f(\xi(t) | \mathbf{Y}(t-1))}{\int f(\mathbf{y}(t) | \tilde{\xi}(t)) f(\tilde{\xi}(t) | \mathbf{Y}(t-1)) d\tilde{\xi}(t)} \quad (49)$$

Note that each target has an associated real-valued amplitude  $x^{(n)}(t)$  and a location on a large discrete grid  $s^{(n)}(t) \in \{1, 2, \dots, Q\}$  for large  $Q$ . Thus, the joint densities  $f(\xi(t) | \mathbf{Y}(t-1))$  and  $f(\xi(t) | \mathbf{Y}(t-1))$  are in general very high-dimensional functions that may be intractable to estimate exactly. Under certain assumptions, however, it may be possible to derive exact equations for these updates.

#### B. Static case

In the static case when  $\alpha = \beta = 0$  and  $\pi_0 = 1$ , we have the simple situation where

$$\mathbf{S}(t) = \mathbf{S}(t-1). \quad (50)$$

Since targets are fixed in position and cannot occupy the same cell by Assumption 7, we can easily show that the joint density factors into:

$$f(\xi(t) | \mathbf{Y}(t')) = f(\Psi(t), \Theta(t) | \mathbf{Y}(t')) = f(\Theta(t) | \mathbf{Y}(t'), \Psi(t)) f(\Psi(t) | \mathbf{Y}(t')) \quad (51)$$

for  $t' = t, t-1$ ,  $\Psi(t) = \{I_i(t)\}_{i=1}^Q$ , and  $\Theta(t) = \{\theta_i(t)\}_{i=1}^Q$ . Moreover,  $\Psi(t) = \Psi$  and since the targets are independent across cells, we have:

$$f(\Theta(t) | \mathbf{Y}(t'), \Psi(t)) = \prod_{i=1}^Q f(\theta_i(t) | \mathbf{y}_i(t'), I_i) \quad (52)$$

$$f(\Psi(t) | \mathbf{Y}(t')) = \prod_{i=1}^Q f(I_i | \mathbf{y}_i(t')) \quad (53)$$

where  $\mathbf{y}_i(t) = \{y_i(t_0)\}_{t_0=1}^t$ . Note that  $\theta_i(t)$  is only defined if  $I_i = 1$ . Conditioned on this event, we furthermore note that  $\theta_i(t)$  and  $y_i(t)|\theta_i(t)$  are normally distributed given the allocations  $\lambda_i(t)$ . Thus, the posteriors  $f(\theta_i(t)|\mathbf{y}_i(t'), I_i)$  for  $t' = t, t-1$  can be updated exactly through the Kalman filter equations:

$$\delta_i(t) = y_i(t) - \sqrt{\lambda_i(t)}\hat{\theta}_i(t|t-1) \quad (54)$$

$$s_i(t) = \lambda_i(t)\hat{\sigma}_i^2(t|t-1) + \sigma^2 \quad (55)$$

$$\Gamma_i(t) = \frac{\hat{\sigma}_i^2(t|t-1)\sqrt{\lambda_i(t)}}{s_i(t)} \quad (56)$$

$$\hat{\theta}_i(t|t) = \hat{\theta}_i(t|t-1) + \Gamma_i(t)\delta_i(t), \quad (57)$$

$$\hat{\sigma}_i^2(t|t) = [1 - \Gamma_i(t)\sqrt{\lambda_i(t)}]\hat{\sigma}_i^2(t|t-1), \quad (58)$$

where  $\delta_i(t)$  is the residual measurement error,  $s_i(t)$  is the update measurement error,  $\Gamma_i(t)$  is the Kalman gain, and  $(\hat{\theta}_i(t|t), \hat{\sigma}_i^2(t|t))$  are the updated state estimates. The predict equations are given by:

$$\hat{\theta}_i(t|t-1) = \hat{\theta}_i(t-1|t-1), \quad (59)$$

$$\hat{\sigma}_i^2(t|t-1) = \hat{\sigma}_i^2(t-1|t-1) + \Delta_\theta^2. \quad (60)$$

Moreover, the posteriors on the indicator functions can be easily computed recursively as

$$f(I_i = 1|\mathbf{y}_i(t)) = \frac{f(y_i(t)|I_i = 1, \mathbf{y}_i(t-1))f(I_i = 1|\mathbf{y}_i(t-1))}{\sum_{j=0,1} f(y_i(t)|I_i = j, \mathbf{y}_i(t-1))f(I_i = j|\mathbf{y}_i(t-1))}, \quad (61)$$

where we note that when  $I_i = 0$

$$y_i(t)|I_i = 0, \mathbf{y}_i(t-1) \sim \mathcal{N}(0, \sigma^2) \quad (62)$$

and when  $I_i = 1$

$$\begin{aligned} f(y_i(t)|I_i = 1, \mathbf{y}_i(t-1)) &= \int f(y_i(t)|\theta_i(t), I_i = 1)f(\theta_i(t)|I_i = 1, \mathbf{y}_i(t-1))d\theta_i(t) \\ &= \int \phi(y_i(t); \sqrt{\lambda_i(t)}\hat{\theta}_i(t|t-1), \sigma^2)\phi(\theta_i(t); \hat{\theta}_i(t|t-1), \hat{\sigma}_i^2(t|t-1))d\theta_i(t) \\ &= \phi(y_i(t); \sqrt{\lambda_i(t)}\hat{\theta}_i(t|t-1), \lambda_i(t)\hat{\sigma}_i^2(t|t-1) + \sigma^2) \end{aligned} \quad (63)$$

where  $\phi(x; \mu, \sigma^2)$  is the Gaussian pdf with mean  $\mu$  and variance  $\sigma^2$  evaluated at  $x$ . From this equation we see that

$$y_i(t)|I_i = 1, \mathbf{y}_i(t-1) \sim \mathcal{N}(\sqrt{\lambda_i(t)}\hat{\theta}_i(t|t-1), \lambda_i(t)\hat{\sigma}_i^2(t|t-1) + \sigma^2) \quad (64)$$

In the static case, we see that updating the posteriors for  $I_i$  for all  $t = 1, 2, \dots, T$  involves (a) updating the conditional mean and variances for  $\theta_i(t)$  given the measurements, and (b) updating the posterior probability for  $I_i = 1$ . This gives insight into an approximate method that we will use in the general case when the targets are allowed to move, enter, or leave the scene.

### C. Approximations in the general case

Similar to the static case, we assume that there are no interacting targets so that we can factor our posterior density into a form that makes it tractable to estimate directly. In order to do this, we use Assumption 8:

**Assumption 8.** There is at most one target in the vicinity of a location

$$|\{n : s^{(n)}(t) \in H(i)\}| \leq 1 \quad (65)$$

for all  $i = 1, 2, \dots, Q$ .

This is clearly more restrictive than Assumption 7. Under this assumption, we have for  $t' = t, t - 1$

$$\begin{aligned} f(\xi(t)|\mathbf{Y}(t')) &= f(\Psi(t), \Theta(t)|\mathbf{Y}(t')) \\ &= f(\Theta(t)|\Psi(t), \mathbf{Y}(t'))f(\Psi(t)|\mathbf{Y}(t')) \\ &= \prod_{i=1}^Q f(\theta_i(t)|\Psi(t), \mathbf{Y}(t'))f(I_i(t)|\mathbf{Y}(t')) \end{aligned} \quad (66)$$

Beginning with the target amplitudes, we note that

$$\begin{aligned} f(\theta_i(t)|\Psi(t), \mathbf{Y}(t-1)) &= \int f(\theta_i(t)|I_i(t) = 1, \Psi_{H(i)}(t-1), \mathbf{Y}(t-1)) \\ &\quad \cdot f(\Psi_{H(i)}(t-1)|I_i(t) = 1, \mathbf{Y}(t-1))d\Psi_{H(i)(t-1)}, \end{aligned} \quad (67)$$

where  $\Psi_{H(i)}(t) = \{I_j(t)\}_{j \in H(i)}$ . Define  $E_{i,j}(t)$  to be the event that assigns  $\Psi_{H(i)}(t)$  as

$$E_{i,j}(t) \triangleq \begin{cases} I_k(t) = 1, & j = k, \\ I_k(t) = 0, & j \neq k, \end{cases} \quad \forall k \in H(i) \quad (68)$$

The event  $E_{i,0}(t)$  refers to the case where a target is added to the scene at location  $i$  at time  $t$ . Then,

under the assumption that at most one target exists in the vicinity of a cell, we have

$$\begin{aligned}
& f(\theta_i(t)|\Psi(t), \mathbf{Y}(t-1)) \\
&= \sum_{j \in H(i) \cup \{0\}} f(\theta_i(t)|I_i(t) = 1, E_{i,j}(t-1), \mathbf{Y}(t-1)) \\
&\quad \cdot f(E_{i,j}(t-1)|I_i(t) = 1, \mathbf{Y}(t-1)) \\
&= \sum_{j \in H(i) \cup \{0\}} \int f(\theta_i(t)|\theta_j(t-1)) \\
&\quad \cdot f(E_{i,j}(t-1)|I_i(t) = 1, \mathbf{Y}(t-1)) \\
&\quad \cdot f(\theta_j(t-1)|I_j(t-1) = 1, \mathbf{Y}(t-1)) d\theta_j(t-1),
\end{aligned} \tag{69}$$

where it is understood that in the case where a new target is added to the scene

$$f(\theta_0(t-1)|I_0(t-1) = 1, \mathbf{Y}(t-1)) = f(\theta_0(t-1)) \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2) \tag{70}$$

and

$$\theta_i(t) = \theta_0(t-1) \tag{71}$$

In the static case, both  $f(\theta_i(t)|\theta_j(t-1))$  and  $f(\theta_j(t-1)|I_j(t-1) = 1, \mathbf{Y}(t-1))$  are Gaussian which makes it possible to analytically integrate equation (69). Indeed, at time  $t = 1$ , it can be easily seen that  $\theta_j(1) \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ . However, for  $t > 1$ , equation (69) shows that we get a Gaussian mixture model with mixing coefficients given by

$$f(E_{i,j}(t-1)|I_i(t) = 1, \mathbf{Y}(t-1)) \tag{72}$$

In order to make the estimation of the posterior distributions very simple, we make the assumption that

$$f(E_{i,j}(t-1)|I_i(t) = 1, \mathbf{Y}(t-1)) = \mathbf{1}_{\{j=j^*\}} \tag{73}$$

for a single  $j^* \in H(i) \cup \{0\}$ . In other words, conditioned on the event that a target exists at cell  $i$ , it is known with probability 1 that the target transitioned from either a single neighboring cell or entered the scene at time  $t$ . This assumption is restrictive except at high SNR. However, it allows us to simplify equation (69) as

$$\begin{aligned}
f(\theta_i(t)|\Psi(t), \mathbf{Y}(t-1)) &= \int f(\theta_i(t)|\theta_{j^*}(t-1)) \\
&\quad \cdot f(\theta_{j^*}(t-1)|I_{j^*}(t-1) = 1, \mathbf{Y}(t-1)) d\theta_{j^*}(t-1),
\end{aligned} \tag{74}$$

which can easily be seen to be Gaussian distributed as long as  $f(\theta_j(t)|I_j(t) = 1, \mathbf{Y}(t))$  is Gaussian. Indeed, we see the recursion

$$f(\theta_i(t)|I_i(t) = 1, \mathbf{Y}(t)) \propto f(y_i(t)|I_i(t), \theta_i(t)) f(\theta_i(t)|I_i(t) = 1, \mathbf{Y}(t-1)) \tag{75}$$

Using equations (74) and (75), it is simple to show that a simply modified Kalman filter will give the exact recursion required to update the posterior densities. In fact, it is the same recursion given in the static case, except that we have

$$\hat{\theta}_i(t|t-1) = \hat{\theta}_{j_i^*(t-1)}(t-1|t-1), \quad (76)$$

$$\hat{\sigma}_i^2(t|t-1) = \hat{\sigma}_{j_i^*(t-1)}^2(t-1|t-1) + \Delta_\theta^2 \quad (77)$$

$$j_i^*(t-1) = \arg \max_{j \in H(i) \cup \{0\}} f(E_{i,j}(t-1) | I_i(t) = 1, \mathbf{Y}(t-1)). \quad (78)$$

**Proposition 6.** *When (73) holds for all  $i \in \Psi(t)$  and for all  $t = 1, 2, \dots, T$ , then the cost function (10) is proportional to the minimum weighted mean squared error*

$$J_T(\boldsymbol{\lambda}) = \mathbb{E} \left[ \sum_{t=1}^T \gamma(t) \sum_{i=1}^Q I_i(t) \left( \theta_i(t) - \hat{\theta}_i(t) \right)^2 \right],$$

where  $\hat{\theta}_i(t) = \mathbb{E} \left\{ \theta_i(t) \middle| I_i(t) = 1, \mathbf{Y}(t) \right\}$ .

*Proof:* We first note that

$$\begin{aligned} \mathbb{E} \left[ \sum_{i=1}^Q I_i(t) \left( \theta_i(t) - \hat{\theta}_i(t) \right)^2 \right] &= \sum_{i=1}^Q I_i(t) \text{var} \left\{ \theta_i(t) \middle| I_i(t) = 1, \mathbf{Y}(t) \right\}, \\ &= \sum_{i=1}^Q I_i(t) \hat{\sigma}_{j_i^*(t-1)}^2(t|t) \end{aligned} \quad (79)$$

since  $\hat{\theta}_i(t)$  is the conditional mean estimator by definition and using the definitions in (73) and (78).

Moreover, by the update equations (54)-(58), it can easily be shown that

$$\hat{\sigma}_i^2(t|t) = \sigma^2 \left[ \frac{\sigma^2}{\hat{\sigma}_i^2(t|t-1)} + \lambda_i(t) \right]^{-1}. \quad (80)$$

Thus we have

$$\begin{aligned} \mathbb{E} \left[ \sum_{t=1}^T \gamma(t) \sum_{i=1}^Q I_i(t) \left( \theta_i(t) - \hat{\theta}_i(t) \right)^2 \right] &= \mathbb{E} \left[ \sum_{t=1}^T \gamma(t) \sum_{i=1}^Q I_i(t) \hat{\sigma}_{j_i^*(t-1)}^2(t|t) \right] \\ &\propto \mathbb{E} \left[ \sum_{t=1}^T \gamma(t) \sum_{i=1}^Q I_i(t) \left[ \frac{\sigma^2}{\hat{\sigma}_i^2(t|t-1)} + \lambda_i(t) \right]^{-1} \right] \\ &= \mathbb{E} \left[ \sum_{t=1}^T \gamma(t) \sum_{i=1}^Q \frac{p_i(t)}{\sigma^2 / \hat{\sigma}_i^2(t|t-1) + \lambda_i(t)} \right] = J_T(\boldsymbol{\lambda}), \end{aligned} \quad (81)$$

where the last equality occurs because  $\mathbb{E} \{ I_i(t) | \mathbf{Y}(t) \} = p_i(t)$ . ■

Looking at the update equations for the target indicators, we get

$$f(I_i(t) = 1 | \mathbf{Y}(t-1)) = \sum_{j \in H(i) \cup \{0\}} f(I_i(t) = 1 | E_{i,j}(t-1)) f(E_{i,j}(t-1) | \mathbf{Y}(t-1)) \quad (82)$$

and

$$f(I_i(t) = 1|\mathbf{Y}(t)) \propto f(y_i(t)|I_i(t) = 1, \mathbf{Y}(t-1))f(I_i(t) = 1|\mathbf{Y}(t-1)), \quad (83)$$

where

$$\begin{aligned} & f(y_i(t)|I_i(t) = 1, \mathbf{Y}(t-1)) \\ &= \int f(y_i(t)|I_i(t) = 1, \theta_i(t)\mathbf{Y}(t-1)) \\ & \quad \cdot f(\theta_i(t)|I_i(t) = 1, \mathbf{Y}(t-1))d\theta_i(t) \end{aligned} \quad (84)$$

Similar to the derivation in the static case, it can easily be seen that

$$y_i(t)|I_i(t) = 1, \mathbf{Y}(t-1) \sim \mathcal{N}(\sqrt{\lambda_i(t)}\mu_i(t|t-1), \lambda_i(t)\sigma_i^2(t|t-1) + \sigma^2) \quad (85)$$

and

$$y_i(t)|I_i(t) = 0, \mathbf{Y}(t-1) \sim \mathcal{N}(0, \sigma^2) \quad (86)$$

#### *D. Discussion of generalizations of state model and posterior estimation methods*

As mentioned earlier, it is a difficult, if not intractable, problem to exactly estimate the posterior distribution of  $\xi(t)$  given  $\mathbf{Y}(t-1)$  that is required for our adaptive algorithms. We have provided a simple algorithm that approximates the posterior distribution under some restrictive assumptions. A simple way to alleviate these restrictions is to use a particle filter implementation for  $\xi(t)$  or other approximate method (e.g., the extended and unscented Kalman filters).

Moreover, we have provided a particular state model that builds on our previous work with the inclusion of transition, birth, and death probabilities. However, there are many other models for dynamic state models, including linear and nonlinear motion models, targets that may occupy multiple adjacent cells, and various noise models. In any of these cases, one would have to use a different posterior estimation algorithm to provide estimates of  $\Pr(I_i(t) = 1|\mathbf{Y}(t-1))$ .

Future work plans to compare other posterior estimation algorithms such as the JMDF particle filter [16] to the one presented in this work, as well as generalizations to more interesting dynamic state models.

#### *E. Unobservable targets*

One particular generalization of the measurement model that is used in this work is the inclusion of indicator variables for observable/unobservable targets. In many applications, certain locations may be

observed for short durations, such as locations in the null of a radar beam. Define

$$O_i(t) = \begin{cases} 1, & \text{Location } i \text{ is observable} \\ 0, & \text{Otherwise} \end{cases} \quad \text{for } i \in \{1, 2, \dots, Q\}. \quad (87)$$

to be an indicator variable for the observability of the  $i$ -th location. Then the measurement model becomes

$$y_i(t) = \sqrt{\lambda_i(t)} I_i(t) O_i(t) \theta_i(t) + \varepsilon_i(t), \quad (88)$$

It is assumed that  $\mathbf{O} = \{O_i(t)\}_{i,t}$  is known to the user a priori. Thus, we are required to estimate the densities:

$$f(\xi(t) | \mathbf{Y}(t'), \mathbf{O}) \quad (89)$$

for  $t' = t, t - 1$ . We make the simplifying assumption that if  $I_i(t - 1) = 1$ , then

$$O_i(t) = 1 \Leftrightarrow O_j(t) = 1, \forall j \in H(i) \quad (90)$$

It can easily be seen that when  $O_i(t) = 1$ , we have the identical update equations to the fully observable case. However, when  $O_i(t) = 0$ , the predict equations remain the same as before, but the update equations are changed in the following manner:

$$f(I_i(t) = 1 | \mathbf{Y}(t), O_i(t) = 0) = f(I_i(t) = 1 | \mathbf{Y}(t - 1)) \quad (91)$$

and the target amplitudes when  $O_i(t) = 0$ :

$$\hat{\theta}_i(t|t) = \hat{\theta}_i(t|t - 1) \quad (92)$$

$$\hat{\sigma}_i^2(t|t) = \hat{\sigma}_i^2(t|t - 1) \quad (93)$$

## APPENDIX II

### PROOF OF PROPOSITION 1

Conditioned on the number of targets  $|\Psi(1)|$ , the per-stage cost of the omniscient policy can be determined by setting  $p_i(t) = I_i(t)$  and substituting (28) and (29) into (11), yielding

$$M_t(\boldsymbol{\lambda}^o) = \frac{|\Psi(1)|^2}{|\Psi(1)|(\sigma^2/\sigma_\theta^2) + \bar{\Lambda}(t)}.$$

As a function of  $|\Psi(1)|$ ,  $M_t(\lambda^o)$  is proportional to  $h_0(z) = z^2/(z + \bar{\Lambda}(t))$ , where  $z = (\sigma^2/\sigma_\theta^2)|\Psi(1)|$ .

The derivatives of the function  $h_0$  are given by

$$h'_0(z) = \frac{z(z + 2\bar{\Lambda}(t))}{(z + \bar{\Lambda}(t))^2} \quad (94a)$$

$$h''_0(z) = \frac{2\bar{\Lambda}(t)^2}{(z + \bar{\Lambda}(t))^3} \quad (94b)$$

$$h'''_0(z) = -\frac{6\bar{\Lambda}(t)^2}{(z + \bar{\Lambda}(t))^4} \quad (94c)$$

$$h^{(4)}_0(z) = \frac{24\bar{\Lambda}(t)^2}{(z + \bar{\Lambda}(t))^5}, \quad (94d)$$

showing that  $h_0$  has a positive fourth derivative for  $z \geq 0$ . Hence by Taylor's theorem,  $h_0$  is lower bounded by its third-order expansion,

$$h_0(z) \geq h_0(z_0) + h'_0(z_0)(z - z_0) + \frac{h''_0(z_0)}{2}(z - z_0)^2 + \frac{h'''_0(z_0)}{3!}(z - z_0)^3.$$

Letting the center of expansion  $z_0 = \mathbb{E}[z]$  and taking expectations with respect to  $z$ , we obtain

$$\mathbb{E}[h_0(z)] \geq h_0(\mathbb{E}[z]) + \frac{h''_0(\mathbb{E}[z])}{2}\text{var}(z) + \frac{h'''_0(\mathbb{E}[z])}{3!}\mathbb{E}[(z - \mathbb{E}[z])^3]. \quad (95)$$

Noting that  $|\Psi(1)| \sim \text{Binomial}(Q, p_0)$ , the moments of  $z$  are

$$\mathbb{E}[z] = (\sigma^2/\sigma_\theta^2)p_0Q \quad (96a)$$

$$\text{var}(z) = (\sigma^2/\sigma_\theta^2)^2 p_0Q(1 - p_0) \quad (96b)$$

$$\mathbb{E}[(z - \mathbb{E}[z])^3] = (\sigma^2/\sigma_\theta^2)^3 p_0Q(1 - p_0)(1 - 2p_0). \quad (96c)$$

Substituting (94) and (96) into (95) and using the definition  $r_0(t) = \sigma^2Q/(\sigma_\theta^2\bar{\Lambda}(t))$  to simplify, we obtain

$$\mathbb{E}\{M_t(\lambda^o)\} \geq \frac{(p_0Q)^2}{\bar{\Lambda}(t)(1 + p_0r_0(t))} + \frac{p_0Q(1 - p_0)}{\bar{\Lambda}(t)(1 + p_0r_0(t))^3} - \frac{(1 - p_0)(1 - 2p_0)p_0r_0(t)}{\bar{\Lambda}(t)(1 + p_0r_0(t))^4}.$$

This bound holds for each  $t$  and therefore the total cost is bounded as

$$J_T(\lambda^o) \geq \sum_{t=1}^T \frac{\gamma(t)}{\bar{\Lambda}(t)} \left[ \frac{(p_0Q)^2}{1 + p_0r_0(t)} + \frac{p_0Q(1 - p_0)}{(1 + p_0r_0(t))^3} - \frac{(1 - p_0)(1 - 2p_0)p_0r_0(t)}{(1 + p_0r_0(t))^4} \right]. \quad (97)$$

For the uniform allocation policy defined by  $\lambda_i^u(t) = \Lambda(t)/Q$  for all  $i$ , a formula similar to (29) holds for  $c_i(t)$  and the per-stage cost (11) is

$$M_t(\lambda^u) = \frac{Q|\Psi(1)|}{Q(\sigma^2/\sigma_\theta^2) + \bar{\Lambda}(t)} = \frac{Q|\Psi(1)|}{\bar{\Lambda}(t)(1 + r_0(t))}, \quad (98)$$

again conditioned on  $|\Psi(1)|$ . Hence the total cost is

$$J_T(\lambda^u) = \sum_{t=1}^T \frac{\gamma(t)}{\bar{\Lambda}(t)} \frac{p_0Q^2}{1 + r_0(t)}. \quad (99)$$

The result follows from the ratio of (97) and (99).



### APPENDIX III

#### PROOF OF LEMMA 1

By definition of  $\bar{c}(1)$ , the lemma is true for  $t = 1$ . We proceed by induction, considering first the case  $\bar{c}(t) < c_{\text{crit}}$ . First it is shown that under the assumptions of the lemma and  $\bar{c}(t) < c_{\text{crit}}$ , all locations in  $H(\Psi(t-1))$  receive nonzero allocations in stage  $t$ . Given (31) and Assumptions 4 and 5, the permutation  $\chi$  in (18) ranks all of the indices  $i \in \Psi(t-1)$  in the ROI equally, followed by  $i \in G(\Psi(t-1))$ , again all equally. It is then straightforward to see that the sequence  $g(k)$  (19) that determines the number of nonzero allocations is as follows:

$$g(k) = \begin{cases} 0, & k = 1, \dots, |\Psi(1)| - 1, \\ \bar{c}(t)|\Psi(1)|e(\pi_0, G), & k = |\Psi(1)|, \dots, |H(\Psi(1))| - 1, \\ \infty & k \geq |H(\Psi(1))|. \end{cases} \quad (100)$$

In particular, given Assumption 6 and  $\bar{c}(t) < c_{\text{crit}}$ , we have

$$g(|H(\Psi(1))| - 1) < c_{\text{crit}}|\Psi(1)|e(\pi_0, G) = \Lambda < g(|H(\Psi(1))|),$$

implying that the number of nonzero allocations  $k^* = |H(\Psi(1))|$ .

We now use (27) and the fact that  $\lambda_i(t) > 0$  for  $i \in H(\Psi(t-1))$  to propagate the posterior precisions forward in time. Combining (27) and (20), we have

$$\begin{aligned} c_i(t+1) &= \left( \Lambda + \sum_{j \in H(\Psi(t-1))} c_j(t) \right) \frac{\sqrt{p_{s^{(n)}(t)}(t)}}{\sum_{j \in H(\Psi(t-1))} \sqrt{p_j(t)}} \\ &= \left( \Lambda + \sum_{j \in H(\Psi(t-1))} c_j(t) \right) \frac{\sqrt{p_{s^{(n)}(t)}(t)}}{|\Psi(1)| \left( \sqrt{\pi_0} + \sqrt{|G|(1-\pi_0)} \right)}, \quad i \in H(s^{(n)}(t)), \end{aligned} \quad (101)$$

using (31) in the second equality. Conditioned on  $|\Psi(1)|$ , there are two random quantities in (101): the probability  $p_{s^{(n)}(t)}(t)$ , which depends on target motion between stages  $t-1$  and  $t$ , and the precisions  $c_j(t)$ , which depend on target motion up to stage  $t-1$ . These quantities are independent according to

the target model. Thus taking the conditional expectation of both sides of (101) yields

$$\begin{aligned}\mathbb{E}[c_i(t+1) \mid |\Psi(1)|] &= \left( \Lambda + \sum_{j \in H(\Psi(t-1))} \mathbb{E}[c_j(t) \mid |\Psi(1)|] \right) \frac{\mathbb{E}[\sqrt{p_{s^{(n)}(t)}(t)} \mid |\Psi(1)|]}{|\Psi(1)| (\sqrt{\pi_0} + \sqrt{|G|(1-\pi_0)})} \\ &= \left( \Lambda + \sum_{j \in H(\Psi(t-1))} \mathbb{E}[c_j(t) \mid |\Psi(1)|] \right) \frac{\pi_0^{3/2} + \frac{1}{\sqrt{|G|}}(1-\pi_0)^{3/2}}{|\Psi(1)| (\sqrt{\pi_0} + \sqrt{|G|(1-\pi_0)})} \quad (102a)\end{aligned}$$

$$\leq \left( \frac{\Lambda}{|\Psi(1)|} + (1+|G|)\bar{c}(t) \right) \frac{\pi_0^{3/2} + \frac{1}{\sqrt{|G|}}(1-\pi_0)^{3/2}}{\sqrt{\pi_0} + \sqrt{|G|(1-\pi_0)}} \quad (102b)$$

$$\equiv \bar{c}(t+1), \quad i \in H(s^{(n)}(t)). \quad (102c)$$

The second line (102a) follows from (31) and because  $s^{(n)}(t) = s^{(n)}(t-1)$  with probability  $\pi_0$  and  $s^{(n)}(t) \in G(s^{(n)}(t-1))$  with probability  $1-\pi_0$ . In the third line (102b), we have used the inductive assumption  $\bar{c}(t) \geq \mathbb{E}[c_j(t) \mid |\Psi(1)|]$ ,  $j \in H(\Psi(t-1))$ , and the equality  $|H(\Psi(t-1))| = (1+|G|)|\Psi(1)|$ . The last line (102c) follows from the recursion (32), thus completing the induction.

Next we consider the case  $\bar{c}(t) \geq c_{\text{crit}}$ , using induction as before. The base case, i.e., the first  $t$  such that  $\bar{c}(t) \geq c_{\text{crit}}$ , is either true for  $t=1$  or follows eventually from the previous induction for  $\bar{c}(t) < c_{\text{crit}}$ . Similar to above, it can be shown that under Assumptions 4–6 and  $\bar{c}(t) \geq c_{\text{crit}}$ ,

$$0 = g(|\Psi(1)| - 1) < \Lambda \leq g(|\Psi(1)|),$$

implying that only the previous target locations  $i \in \Psi(t-1)$  are allocated nonzero effort in stage  $t$ . In the case  $s^{(n)}(t) = s^{(n)}(t-1)$ , which occurs with probability  $\pi_0$ , we combine (27) with (20) to obtain

$$c_i(t+1) = \frac{1}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} c_j(t) \right), \quad i \in H(s^{(n)}(t)).$$

In the other case  $s^{(n)}(t) \in G(s^{(n)}(t-1))$  with probability  $1-\pi_0$ ,  $\lambda_{s^{(n)}(t)}(t) = 0$  and  $c_i(t+1) = c_{s^{(n)}(t)}(t)$ . Therefore the expected precision conditioned on  $|\Psi(1)|$  is given by

$$\begin{aligned}\mathbb{E}[c_i(t+1) \mid |\Psi(1)|] &= \frac{\pi_0}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} \mathbb{E}[c_j(t) \mid |\Psi(1)|] \right) + (1-\pi_0)\mathbb{E}[c_{s^{(n)}(t)}(t) \mid |\Psi(1)|] \\ &\leq \bar{c}(t) + \frac{\pi_0 \Lambda}{|\Psi(1)|} \\ &= \bar{c}(t+1), \quad i \in H(s^{(n)}(t)),\end{aligned}$$

using the inductive assumption on  $\bar{c}(t)$  and (32) to complete the proof.

## APPENDIX IV

## PROOF OF PROPOSITION 2

We rewrite the expected per-stage cost by combining (12) and (27) and iterating expectations to yield

$$\mathbb{E}[M_t(\boldsymbol{\lambda})] = \mathbb{E} \left\{ \sum_{n=1}^{|\Psi(1)|} \mathbb{E} \left[ \frac{1}{c_{s^{(n)}(t+1)}(t+1)} \mid |\Psi(1)| \right] \right\}.$$

Using the convexity of the function  $1/x$  and Jensen's inequality, this may be bounded from below as

$$\mathbb{E}[M_t(\boldsymbol{\lambda})] \geq \mathbb{E} \left\{ \sum_{n=1}^{|\Psi(1)|} \frac{1}{\mathbb{E}[c_{s^{(n)}(t+1)}(t+1) \mid |\Psi(1)|]} \right\}.$$

For  $t$  large enough such that  $\bar{c}(t) \geq c_{\text{crit}}$ , Lemma 1 applies to provide a further lower bound,

$$\mathbb{E}[M_t(\boldsymbol{\lambda})] \geq \mathbb{E} \left\{ \frac{|\Psi(1)|}{\bar{c}(t+1)} \right\}. \quad (103)$$

The recursion (32) for  $\bar{c}(t)$  implies that

$$\bar{c}(t) = \frac{\pi_0 \Lambda}{|\Psi(1)|} t + O(1). \quad (104)$$

Substituting (104) into (103) and recalling that  $|\Psi(1)|$  is binomially distributed with parameters  $Q$  and  $p_0$ , we obtain

$$\begin{aligned} \mathbb{E}[M_t(\boldsymbol{\lambda}^s)] &\geq \mathbb{E} \left\{ \frac{|\Psi(1)|^2}{\pi_0 \Lambda t} + O\left(\frac{1}{t^2}\right) \right\} \\ &= \frac{p_0 Q (p_0 Q + 1 - p_0)}{\pi_0 \Lambda t} + O\left(\frac{1}{t^2}\right). \end{aligned}$$

## APPENDIX V

## PROOF OF PROPOSITION 3

First we derive an expression for the steady-state (large  $t$ ) posterior variance. For the uniform and omniscient policies under Assumption 6, the effort allocation  $\lambda_{s^{(n)}(t)}(t)$  for targets is independent of both  $n$  and  $t$ . Therefore all targets have the same steady-state posterior variance  $\sigma_{\text{ss}}^2$ , which may be determined by setting  $i = s^{(n)}(t+1)$  and  $\sigma_{s^{(n)}(t+1)}^2(t+1) = \sigma_{s^{(n)}(t)}^2(t) = \sigma_{\text{ss}}^2$  in (26) to yield the following quadratic equation:

$$\lambda_{s^{(n)}(t)}(t) (\sigma_{\text{ss}}^4 - \Delta^2 \sigma_{\text{ss}}^2) - \sigma^2 \Delta^2 = 0. \quad (105)$$

Taking the positive root results in

$$\sigma_{\text{ss}}^2 = \frac{\Delta^2}{2} \left( 1 + \sqrt{1 + \frac{4\sigma^2}{\Delta^2 \lambda_{s^{(n)}(t)}(t)}} \right). \quad (106)$$

To relate the steady-state variance (106) to the gain (30), we take  $T \rightarrow \infty$  and use Assumption 3, which reduces the gain to a ratio of steady-state expected per-stage costs. To compute the per-stage cost, we first note that the denominator  $c_i(t) + \lambda_i(t)$  in (12) corresponds to the posterior variance after measurement but before the increment  $\Delta^2$ , while  $\sigma_{ss}^2$  in (106) is the steady-state variance after the increment. Hence the steady-state per-stage cost conditioned on  $\Psi(t)$  is

$$\begin{aligned} \lim_{t \rightarrow \infty} \sum_{i \in \Psi(t)} \frac{1}{c_i(t) + \lambda_i(t)} &= |\Psi(1)| \frac{\sigma_{ss}^2 - \Delta^2}{\sigma^2} \\ &= \frac{\Delta^2 |\Psi(1)|}{2\sigma^2} \left( \sqrt{1 + \frac{4\sigma^2}{\Delta^2 \lambda_{s^{(n)}(t)}(t)}} - 1 \right). \end{aligned} \quad (107)$$

For the uniform policy,  $\lambda_{s^{(n)}(t)}(t) = \Lambda/Q$  and the expectation over  $|\Psi(1)|$  gives

$$\lim_{t \rightarrow \infty} \mathbb{E} \{M_t(\boldsymbol{\lambda}^u)\} = \frac{\Delta^2 p_0 Q}{2\sigma^2} \left( \sqrt{1 + 4r_+} - 1 \right). \quad (108)$$

For the omniscient policy,  $\lambda_{s^{(n)}(t)}(t) = \Lambda/|\Psi(1)|$  and (107) is proportional to  $h_+(z) = z(\sqrt{1+z} - 1)$ , where  $z = 4\sigma^2 |\Psi(1)| / (\Delta^2 \Lambda)$ . The function  $h_+$  has derivatives

$$h'_+(z) = \sqrt{1+z} - 1 + \frac{z}{2\sqrt{1+z}} \quad (109a)$$

$$h''_+(z) = \frac{1 + 3z/4}{(1+z)^{3/2}} \quad (109b)$$

$$h'''_+(z) = -\frac{3(2+z)}{8(1+z)^{5/2}} \quad (109c)$$

$$h^{(4)}_+(z) = \frac{3(8+3z)}{16(1+z)^{7/2}}. \quad (109d)$$

Hence similar to  $h_0$  in the proof of Proposition 1,  $h_+$  has a positive fourth derivative and the same lower bound (95) applies to  $\mathbb{E}[h_+(z)]$ . Since  $z$  is again proportional to  $|\Psi(1)|$ , the moments of  $z$  are given by similar expressions as in (96). Combining these moments with (95), (109), and the definition of  $r_+$  yields after some simplification

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbb{E} \{M_t(\boldsymbol{\lambda}^o)\} &\geq \frac{\Delta^2 p_0 Q}{2\sigma^2} \left( \sqrt{1 + 4p_0 r_+} - 1 \right) + \frac{p_0 Q(1-p_0)}{\Lambda} \frac{1 + 3p_0 r_+}{(1 + 4p_0 r_+)^{3/2}} \\ &\quad - \frac{p_0(1-p_0)(1-2p_0)}{\Lambda} \frac{r_+(1 + 2p_0 r_+)}{(1 + 4p_0 r_+)^{5/2}}. \end{aligned} \quad (110)$$

Taking the ratio of (110) and (108) yields the result.

## APPENDIX VI

### PROOF OF PROPOSITION 4

As in Proposition 3, under Assumption 3 the gain reduces to the ratio of steady-state expected per-stage costs. To compute the per-stage cost for the semi-omniscient policy, we first show that the precisions

$c_i(t)$  are small as claimed. Rewriting the evolution equation (26) in terms of  $c_i(t)$  gives

$$c_i(t+1) = \frac{\sigma^2}{\Delta^2} \frac{c_{s^{(n)}(t)}(t) + \lambda_{s^{(n)}(t)}(t)}{c_{s^{(n)}(t)}(t) + \lambda_{s^{(n)}(t)}(t) + \sigma^2/\Delta^2} < \frac{\sigma^2}{\Delta^2}, \quad i \in H(s^{(n)}(t)). \quad (111)$$

This bound together with assumption (35) imply that the semi-omniscient policy allocates nonzero effort to all locations in  $H(\Psi(t-1))$ , i.e., the sequence  $g(k)$  (19) satisfies  $g(k) < \Lambda$  for  $k = |H(\Psi(t-1))| - 1$ . Using (31), the sum of probability ratios  $\sqrt{p_{\chi(i)}(t)/p_{\chi(k+1)}(t)}$  in (19) can be bounded by  $Q(e(\pi_0, G) + 1)$ . Combining this with (111), (35), and the definition of  $r_+$  yields

$$g(|H(\Psi(t-1))| - 1) < \frac{\sigma^2 Q}{\Delta^2} (e(\pi_0, G) + 1) \leq \Lambda$$

as desired.

Given that  $\lambda_i(t) > 0$  for all  $i \in H(\Psi(t-1))$ , the per-stage cost for the semi-omniscient policy can be computed from (11), (20) and (31) as

$$\begin{aligned} M_t(\boldsymbol{\lambda}^s) &= \frac{\left(\sum_{i \in H(\Psi(t-1))} \sqrt{p_i(t)}\right)^2}{\Lambda + \sum_{i \in H(\Psi(t-1))} c_i(t)} \\ &\geq \frac{\left(\sum_{i \in H(\Psi(t-1))} \sqrt{p_i(t)}\right)^2}{\Lambda + (1 + |G|)|\Psi(1)|\sigma^2/\Delta^2} \\ &= \frac{|\Psi(1)|^2 \left(\sqrt{\pi_0} + \sqrt{|G|(1 - \pi_0)}\right)^2}{\Lambda + (1 + |G|)|\Psi(1)|\sigma^2/\Delta^2} \end{aligned} \quad (112)$$

where the inequality follows from (111) and  $|H(\Psi(t-1))| = (1 + |G|)|\Psi(1)|$ . As a function of  $|\Psi(1)|$ , the right-hand side of (112) has the same form as the function  $h_0(z)$  in the proof of Proposition 1 in Appendix II. Applying the same technique as before and simplifying the resulting expressions, the expectation with respect to  $|\Psi(1)|$  can be bounded as

$$\begin{aligned} \mathbb{E}\{M_t(\boldsymbol{\lambda}^s)\} &\geq \frac{p_0 Q^2}{\Lambda} \left(\sqrt{\pi_0} + \sqrt{|G|(1 - \pi_0)}\right)^2 \\ &\quad \times \left( \frac{p_0}{1 + p_0 q_+} + \frac{1 - p_0}{Q} \frac{1}{(1 + p_0 q_+)^3} - \frac{(1 - p_0)(1 - 2p_0)}{Q^2} \frac{q_+}{(1 + p_0 q_+)^4} \right). \end{aligned} \quad (113)$$

The result is obtained from the ratio of (113) and (108).

APPENDIX VII  
PROOF OF LEMMA 2

As in the proof of Proposition 4, the evolution of the posterior precisions is given by (111). Under Assumptions 4–6 and  $\bar{c}(t) \geq c_{\text{crit}}$ , similar to the proof of Proposition 2 we have

$$c_{s^{(n)}(t)}(t) + \lambda_{s^{(n)}(t)}(t) = \begin{cases} \frac{1}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} c_j(t) \right), & s^{(n)}(t) = s^{(n)}(t-1) \text{ w.p. } \pi_0, \\ c_{s^{(n)}(t)}(t), & s^{(n)}(t) \in G(s^{(n)}(t-1)) \text{ w.p. } 1 - \pi_0. \end{cases}$$

Substituting this into (111) results in

$$\begin{aligned} \mathbb{E}[c_i(t+1) \mid |\Psi(1)|] &= \pi_0 \mathbb{E} \left\{ \frac{(\sigma^2/\Delta^2) \frac{1}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} c_j(t) \right)}{(\sigma^2/\Delta^2) + \frac{1}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} c_j(t) \right)} \mid |\Psi(1)| \right\} \\ &\quad + (1 - \pi_0) \mathbb{E} \left\{ \frac{(\sigma^2/\Delta^2) c_{s^{(n)}(t)}(t)}{(\sigma^2/\Delta^2) + c_{s^{(n)}(t)}(t)} \mid |\Psi(1)| \right\}, \quad i \in H(s^{(n)}(t)). \end{aligned}$$

The terms on the right-hand side are of the form  $ax/(a+x)$  with  $a > 0$ , which is a concave and increasing function of  $x$ . Applying Jensen's inequality and the assumption on  $\bar{c}(t)$ , we obtain

$$\begin{aligned} \mathbb{E}[c_i(t+1) \mid |\Psi(1)|] &\leq \pi_0 \frac{(\sigma^2/\Delta^2) \frac{1}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} \mathbb{E}[c_j(t) \mid |\Psi(1)|] \right)}{(\sigma^2/\Delta^2) + \frac{1}{|\Psi(1)|} \left( \Lambda + \sum_{j \in \Psi(t-1)} \mathbb{E}[c_j(t) \mid |\Psi(1)|] \right)} \\ &\quad + (1 - \pi_0) \frac{(\sigma^2/\Delta^2) \mathbb{E}[c_{s^{(n)}(t)}(t) \mid |\Psi(1)|]}{(\sigma^2/\Delta^2) + \mathbb{E}[c_{s^{(n)}(t)}(t) \mid |\Psi(1)|]} \\ &\leq \pi_0 \frac{(\sigma^2/\Delta^2)(\bar{c}(t) + \Lambda/|\Psi(1)|)}{(\sigma^2/\Delta^2) + \bar{c}(t) + \Lambda/|\Psi(1)|} + (1 - \pi_0) \frac{(\sigma^2/\Delta^2)\bar{c}(t)}{(\sigma^2/\Delta^2) + \bar{c}(t)} \\ &= \bar{c}(t+1), \quad i \in H(s^{(n)}(t)), \end{aligned}$$

as desired.

APPENDIX VIII  
PROOF OF LEMMA 4

First we show that (38) has three distinct real roots. This is equivalent to the discriminant of (38) being positive. Let  $a_1 = \Delta^2/\sigma^2$  and  $a_2 = |\Psi(1)|/\Lambda$ . Noting that (38) lacks a quadratic term, the discriminant can be simplified to

$$D = \pi_0 a_1^3 (4a_1^3 + 12a_1^2 a_2 + (12 - 27\pi_0)a_1 a_2^2 + 4a_2^3). \quad (114)$$

In the omniscient case  $\pi_0 = 1$ , it is known that (38) has three real roots and hence both  $D$  and the quantity in the outer parentheses in (114) are positive. As  $\pi_0$  decreases from 1, the parenthesized quantity only

increases and therefore  $D > 0$  for  $0 < \pi_0 < 1$  as well. Now given that (38) has three real roots, it can be seen that one of the roots is positive and the other two are negative. This is because the coefficients of (38) constrain the product of the roots to be positive and their sum to be zero. The positive root can then be expressed in terms of trigonometric functions as

$$\bar{c}_{\text{ss}}^{-1} = 2\sqrt{\frac{a_1(a_1 + a_2)}{3\pi_0}} \cos\left(\frac{1}{3} \arccos\left(\frac{3\sqrt{3}\pi_0}{2} \frac{a_2}{a_1 + a_2} \sqrt{\frac{a_1}{a_1 + a_2}}\right)\right). \quad (115)$$

We now use the assumption that  $\sqrt{\Delta^2\Lambda/\sigma^2} \ll 1$ , implying that  $\sqrt{a_1/a_2} \ll 1$ , to simplify the expression in (115). First we expand the argument of the arccos function to lowest order in  $a_1/a_2$ :

$$\frac{a_2}{a_1 + a_2} \sqrt{\frac{a_1}{a_1 + a_2}} = \sqrt{\frac{a_1}{a_2}} \left(1 + O\left(\frac{a_1}{a_2}\right)\right).$$

It then follows from further expansions that

$$\arccos\left(\frac{3\sqrt{3}\pi_0}{2} \frac{a_2}{a_1 + a_2} \sqrt{\frac{a_1}{a_1 + a_2}}\right) = \frac{\pi}{2} - \frac{3}{2} \sqrt{\frac{3\pi_0 a_1}{a_2}} + O\left(\frac{a_1}{a_2}\right)$$

and

$$\cos\left(\frac{1}{3} \arccos\left(\frac{3\sqrt{3}\pi_0}{2} \frac{a_2}{a_1 + a_2} \sqrt{\frac{a_1}{a_1 + a_2}}\right)\right) = \frac{\sqrt{3}}{2} + \frac{1}{4} \sqrt{\frac{3\pi_0 a_1}{a_2}} + O\left(\frac{a_1}{a_2}\right).$$

Combining this with

$$2\sqrt{\frac{a_1(a_1 + a_2)}{3\pi_0}} = 2\sqrt{\frac{a_1 a_2}{3\pi_0}} \left(1 + O\left(\frac{a_1}{a_2}\right)\right),$$

we obtain

$$\bar{c}_{\text{ss}}^{-1} = \sqrt{\frac{a_1 a_2}{\pi_0}} \left(1 + O\left(\sqrt{\frac{a_1}{a_2}}\right)\right) = \sqrt{\frac{\Delta^2 |\Psi(1)|}{\pi_0 \sigma^2 \Lambda}} \left(1 + O\left(\sqrt{\frac{\Delta^2 \Lambda}{\sigma^2}}\right)\right).$$

## APPENDIX IX

### PROOF OF PROPOSITION 5

In the regime  $\sqrt{\Delta^2\Lambda/\sigma^2} \ll 1$ , the positive root of the cubic equation (38) is given by (39) in Lemma 4. We verify that  $\bar{c}_{\text{ss}}$  in (39) satisfies  $\bar{c}_{\text{ss}} \geq c_{\text{crit}}$  for  $|\Psi(1)| \geq 1$ . Combined with Lemma 3, this will imply that  $\bar{c}_{\text{ss}}$  is a stationary upper bound on the steady-state precision  $c_{\text{ss}}$ . Substituting (39) and (33) for  $\bar{c}_{\text{ss}}$  and  $c_{\text{crit}}$  and neglecting higher-order terms, the condition  $\bar{c}_{\text{ss}} \geq c_{\text{crit}}$  is equivalent to

$$\sqrt{\frac{\pi_0 \sigma^2 |\Psi(1)|}{\Delta^2 \Lambda}} e(\pi_0, G) \geq 1.$$

The above inequality is most stringent for  $|\Psi(1)| = 1$ , in which case it is ensured by assumption (40).

Hence we conclude that  $\bar{c}_{\text{ss}} \geq c_{\text{ss}}$ .

The remainder of the proof uses arguments from the proofs of Propositions 2 and 3. Under Assumption 3, similar to Proposition 3 it suffices to compute the steady-state expected per-stage costs. As in Proposition 2, the expected per-stage cost can be written as

$$\mathbb{E}[M_t(\boldsymbol{\lambda})] = \mathbb{E} \left\{ \sum_{n=1}^{|\Psi(1)|} \left( \mathbb{E} \left[ \frac{1}{c_{s^{(n)}}(t+1)(t+1)} \mid |\Psi(1)| \right] - \frac{\Delta^2}{\sigma^2} \right) \right\},$$

where the additional term  $\Delta^2/\sigma^2$  reflects the fact that the per-stage cost corresponds to the posterior variance before the increment  $\Delta^2$  whereas  $1/c_{s^{(n)}}(t+1)(t+1)$  includes the increment. Applying Jensen's inequality as before, taking the steady-state limit  $t \rightarrow \infty$  and using the bound  $c_{ss} \leq \bar{c}_{ss}$ , for the semi-omniscient policy we have

$$\lim_{t \rightarrow \infty} \mathbb{E}[M_t(\boldsymbol{\lambda}^s)] \geq \mathbb{E} \left\{ |\Psi(1)| \left( \frac{1}{\bar{c}_{ss}} - \frac{\Delta^2}{\sigma^2} \right) \right\}.$$

Substituting (39) for  $1/\bar{c}_{ss}$  gives

$$\lim_{t \rightarrow \infty} \mathbb{E}[M_t(\boldsymbol{\lambda}^s)] \geq \mathbb{E} \left\{ |\Psi(1)| \sqrt{\frac{\Delta^2 |\Psi(1)|}{\pi_0 \sigma^2 \Lambda}} \left( 1 + O \left( \sqrt{\frac{\Delta^2 \Lambda}{\sigma^2}} \right) \right) \right\},$$

using the approximation  $\sqrt{\Delta^2 \Lambda / \sigma^2} \ll 1$ . Since the right-hand side is proportional to  $|\Psi(1)|^{3/2}$ , a convex function of  $|\Psi(1)| \sim \text{Bin}(Q, p_0)$ , another application of Jensen's inequality yields

$$\lim_{t \rightarrow \infty} \mathbb{E}[M_t(\boldsymbol{\lambda}^s)] \geq p_0 Q \sqrt{\frac{\Delta^2 p_0 Q}{\pi_0 \sigma^2 \Lambda}} \left( 1 + O \left( \sqrt{\frac{\Delta^2 \Lambda}{\sigma^2}} \right) \right). \quad (116)$$

For the uniform policy, the steady-state expected per-stage cost may be approximated from (108) in Proposition 3 as

$$\lim_{t \rightarrow \infty} \mathbb{E}[M_t(\boldsymbol{\lambda}^u)] = p_0 Q \sqrt{\frac{\Delta^2 Q}{\sigma^2 \Lambda}} \left( 1 - O \left( \sqrt{\frac{\Delta^2 \Lambda}{\sigma^2}} \right) \right). \quad (117)$$

The result is given by the ratio of (116) and (117).

## REFERENCES

- [1] B. R. Abidi, N. R. Aragam, Y. Yao, and M. A. Abidi, "Survey and analysis of multimodal sensor planning and integration for wide area surveillance," *ACM Computing Surveys (CSUR)*, vol. 41, no. 1, p. 7, 2008.
- [2] R. Castro, R. Willett, and R. Nowak, "Coarse-to-fine manifold learning," in *Proceedings 2004 International Conference on Acoustics, Speech and Signal Processing*, vol. 3, May 2004, pp. iii-992-5.
- [3] R. Castro, R. Willett, and R. Nowak, "Faster rates in regression via active learning," in *Proceedings of the Neural Information Processing Systems Conference (NIPS) 2005*, Vancouver, Canada, December 2005.
- [4] E. Bashan, R. Raich, and A. O. Hero, III, "Optimal Two-Stage Search for Sparse Targets Using Convex Criteria," *IEEE Transactions on Signal Processing*, vol. 56, pp. 5389-5402, 2008.



- [5] E. Bashan, G. Newstadt, and A. O. Hero, III, "Two-Stage Multi-Scale Search for Sparse Targets," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2331–2341, 2011.
- [6] D. Hitchings and D. A. Castanon, "Adaptive sensing for search with continuous actions and observations," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2010, pp. 7443–7448.
- [7] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6222–6235, 2011.
- [8] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 401–404.
- [9] D. Wei and A. O. Hero, III, "Multistage adaptive estimation of sparse signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 783–796, Oct. 2013.
- [10] M. L. Malloy and R. Nowak, "Sequential testing for sparse recovery," *arXiv preprint arXiv:1212.1801*, 2012.
- [11] M. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *CoRR*, vol. abs/1306.6239, 2013.
- [12] V. Krishnamurthy and R. J. Evans, "Hidden Markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 2893–2908, December 2001.
- [13] E. Chong, C. Kreucher, and A. O. Hero, III, "Monte-Carlo-based partially observable Markov decision process approximations for adaptive sensing," in *Discrete Event Systems, 2008. WODES 2008. 9th International Workshop on*. IEEE, 2008, pp. 173–180.
- [14] D. Wei and A. O. Hero, III, "Performance guarantees for adaptive estimation of sparse signals," Nov. 2013, arXiv:1311.6360.
- [15] G. E. Newstadt, "Adaptive sensing techniques for dynamic target tracking and detection with applications to synthetic aperture radars," Ph.D. dissertation, University of Michigan, May 2013.
- [16] C. Kreucher, K. Kastella, and A. O. H. III, "Multitarget tracking using the joint multitarget probability density," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1396–1414, October 2005.
- [17] D. P. Bertsekas and D. A. Castanon, "Rollout algorithms for stochastic scheduling problems," *Journal of Heuristics*, vol. 5, no. 1, pp. 89–108, 1999.
- [18] G. E. Newstadt, D. L. Wei, and A. O. Hero, III, "Adaptive search for sparse dynamic targets," in *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2013.
- [19] J. N. Ash, "Joint imaging and change detection for robust exploitation in interrupted SAR environments," vol. 8746, 2013, pp. 87 460J–87 460J–9. [Online]. Available: <http://dx.doi.org/10.1117/12.2019019>